REDUCED-ORDER MODELING OF

MULTISCALE TURBULENT CONVECTION:

APPLICATION TO DATA CENTER THERMAL MANAGEMENT

A Dissertation
Presented to
The Academic Faculty

by

Jeffrey D. Rambo

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mechanical Engineering

Georgia Institute of Technology

May 2006

REDUCED-ORDER MODELING OF

MULTISCALE TURBULENT CONVECTION:

APPLICATION TO DATA CENTER THERMAL MANAGEMENT

Approved by:

Dr. Yogendra Joshi, Advisor
School of Mechanical Engineering
Georgia Institute of Technology

Dr. P. K. Yeung
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Shelton M. Jeter
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Benjamin Shapiro
School of Aerospace Engineering
University of Maryland, College Park

Dr. Marc K. Smith
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: February 23, 2006

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Yogendra Joshi, for his support and guidance during my doctoral studies. I am grateful for his insight and career advice, but most of all his confidence in me. I would also like to express my appreciation to my committee members for their valuable suggestions and commitment to this dissertation. I am thankful for all my current and former colleagues for the valuable, or otherwise interesting, discussions we have had concerning this and many other research topics.

I wish to thank all my family and friends who have supported me along the way and above all, I would like to express my deepest gratitude to my loving wife Allison for all her encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

NOMENCLATURE

Symbols

A_{ijk} , C_{ijk} , D_{ij} , D_{ijk} ..... coefficients in Navier-Stokes Galerkin projection

E ..... resolved modal spectrum energy

F_m , F_h ..... mass and heat flux function

G, G_h ..... mass flux and heat flux goal vector

J ..... error functional

Nu ..... Nusselt number

P_k ..... orthogonal subspace projection matrix

Pr_t ..... turbulent Prandtl number

Q ..... heat flux

$R(x,x')$ ..... correlation function

Re ..... Reynolds number

U, T, P ..... velocity and temperature field observations matrices

W ..... optimal / robust solution weight factor

a,b ..... modal weight coefficients

d ..... POD subspace distance

e_i, e_o ..... in-plane and out of plane error components

k ..... turbulent kinetic energy

k_{eff} ..... effective thermal conductivity

m ..... number of observations

n ..... number of system DOF

$\hat{n}$ ..... outward pointing surface normal

p..... number of retained modes in approximation

s..... number of modes in orthogonal complement subspace

$\vec{u}$, T, P..... velocity, temperature and pressure fields

Γ ..... control surface

$\vec{\Phi}$, Ψ, Π ..... velocity, temperature and pressure POD subspace basis vector matrix

Ω, $\partial\Omega$ ..... domain, domain boundary

E..... turbulent kinetic energy dissipation rate

λ..... eigenvalue

ν, ν_{eff} ..... viscosity and effective viscosity

ρ..... density

$\vec{\varphi}$, ψ, π ..... velocity, temperature and pressure POD modes

$\langle\cdot\rangle$ ..... ensemble mean

$\|\cdot\|$ ..... induced or L_2 norm

$\bar{u}$ ..... area-averaged normal velocity

Subscripts

err ..... relative error

0..... source function

Superscripts

$(\cdot)^+$ ..... matrix pseudo-inverse

$(\cdot)^*$ ..... approximate solution or complex conjugate

$(\cdot)^\perp$ ..... orthogonal complement

$(\cdot)^{obs}$ ..... observation quantity

$(\cdot)^T$ ..... matrix transpose

Acronyms

ESV ..... extended state vector

DOF ..... degrees of freedom

FMP ..... flux matching procedure

FNM ..... flow network modeling

GP ..... Galerkin projection

POD ..... proper orthogonal decomposition

PODc ..... orthogonal complement proper orthogonal decomposition

RANS ..... Reynolds-averaged Navier-Stokes

ROM ..... reduced-order model

SUMMARY

Data centers are computing infrastructure facilities used by industries with large data processing needs and the rapid increase in power density of high performance computing equipment has caused many thermal issues in these facilities. Systems-level thermal management requires modeling and analysis of complex fluid flow and heat transfer processes across several decades of length scales. Conventional computational fluid dynamics and heat transfer techniques for such systems are severely limited as a design tool because their large model sizes render parameter sensitivity studies and optimization impractically slow.

The traditional proper orthogonal decomposition (POD) methodology has been reformulated to construct physics-based models of turbulent flows and forced convection. Orthogonal complement POD subspaces were developed to parametrize inhomogeneous boundary conditions and greatly extend the use of the existing POD methodology beyond prototypical flows with fixed parameters. A flux matching procedure was devised to overcome the limitations of Galerkin projection methods for the Reynolds-averaged Navier-Stokes equations and greatly improve the computational efficiency of the approximate solutions. An implicit coupling procedure was developed to link the temperature and velocity fields and further extend the low-dimensional modeling methodology to conjugate forced convection heat transfer. The overall reduced-order modeling framework was able to reduce numerical models containing 10^5 degrees of freedom (DOF) down to less than 20 DOF, while still retaining greater that 90% accuracy over the domain.

Rigorous *a posteriori* error bounds were formulated by using the POD subspace to partition the error contributions and dual residual methods were used to show that the flux matching procedure is a computationally superior approach for low-dimensional modeling of steady turbulent convection.

To efficiently model large-scale systems, individual reduced-order models were coupled using flow network modeling as the component interconnection procedure. The development of handshaking procedures between low-dimensional component models lays the foundation to quickly analyze and optimize the modular systems encountered in electronics thermal management. This modularized approach can also serve as skeletal structure to allow the efficient integration of highly-specialized models across disciplines and significantly advance simulation-based design.

1. INTRODUCTION TO DATA CENTER THERMAL MANAGEMENT

Data centers are computing infrastructure facilities used by industries with large data processing needs, such as telecommunications switching, banking, stock market transactions and supercomputing nodes. The trends of rapidly increasing heat fluxes and volumetric heat generation rates within these devices are causing many unique challenges for their thermal design. Due to the high heating rates, design guidelines followed for human-occupied spaces such as auditoria and theaters are inapplicable for such computing spaces.

The electronics thermal management community has, in the past, focused exclusively on heat removal from the chips and single data processing units or enclosures. The objective has been to remove the heat generated by the chips and associated devices out of the enclosure case and reject it to an extensive ambient. In a data center, hundreds or thousands of these data processing units use the facility as the ambient. Due to the vertical stacking of these components and space constraints, the electronic equipment can interact by the hot exhaust air being drawn from one data processing unit into another unit. Figure 1.1 shows 2 photographs of existing data center facilities.

Detailed experimental measurements and numerical computations of heat transfer and fluid flow are necessary for the proper placement of electronic equipment to ensure its reliable operation. Due to the three-dimensional, multi-mode, multiscale nature of the transport involving at least 10 decades of pertinent length scales from nanometer chip interconnects to the facility length scale of 10's of meters, and prevailing turbulent flow

conditions, such characterization is highly challenging. Only recently is electronics thermal management at the systems level beginning to receive systematic attention. Improved characterization through experimental and computational thermal modeling has the potential of developing energy efficient designs for these facilities, resulting in large financial savings for the end users.

Figure 1.1. Data center photographs, courtesy of W. Tschudi, Lawrence Berkeley National Laboratory

1.1. Power Density Trends

The basic building block of a data center is a rack, which is typically ~2-m tall enclosure, in which various servers, data storage and networking/switching equipment are stacked vertically. In 1990, a typical rack dissipated approximately 1 kW of power [1], while today's racks with the same footprint may dissipate up to 30 kW, based on current server heat loads. The decade from 1992 to 2002 has seen server power density rise 300%, with a projected annual increase of 5% over the next 4 years [2, 3]. The demand for data processing has driven data centers to grow as large as 7,500 m^2 (~80,000 ft^2) [3, 4], producing net power dissipations on the order of several MW. With such an enormous amount of power dissipation, providing an environment for the safe and

reliable operation of the data processing equipment is essential, as these facilities are intended to operate continuously.

Practitioners and data center designers measure data center heat fluxes as the ratio of the net power dissipated to the total footprint of the facility. Recent energy benchmarking studies have shown data centers are operating in the 270 – 800 W/m^2 (25 – 75 W/ft^2) range [3, 4] and growth to 100 W/ft^2 average over the facility, with local regions exceeding 200 W/ft^2, is expected in the near future.

These power densities are well beyond conventional heating, ventilation and air conditioning (HVAC) loads for the cooling of similarly sized rooms such as auditoria and office spaces, which are typically 4 – 8 W/ft^2. A rack footprint is approximately 0.76 m by 0.91 m (2.5 x 3 ft), similar to the area of an auditorium seat. However, a 30 kW rack has 300 times the power dissipation of a sitting person. In other words, applying standard HVAC guidelines to data centers is equivalent to cooling an auditorium with 300 people per seat. With facilities growing up to 5000 m^2 (~50,000 ft^2), the net power dissipated by the data processing equipment could be as large as several MW. The cost of just powering these large computing facilities could be millions of dollars a year, with the cost of providing adequate cooling not far behind.

1.2. Airflow Configurations

A majority of today's data centers use computer room air conditioning (CRAC) units to supply the facility with cooling air designed to provide an adequately low ambient temperature for reliable server operation. A major improvement in airflow management has been to arrange the racks in rows with alternating direction of airflow

[5] such that rack inlets are facing each other, to form a 'cold aisle'. The exhausts of the racks necessarily face each other and form a 'hot' aisle. This hot aisle – cold aisle approach attempts to separate the supply from the exhaust air and increase the overall efficiency of the air delivery and collection from each rack in the data center.

The predominate cooling scheme in today's data centers is to use the CRAC units to supply a raised floor plenum underneath the racks with cold air. Perforated tiles are then located in the cold aisles on the raised floor near the racks to deliver the cool supply air to the data processing equipment. The hot exhaust air from the racks is then collected from the upper portion of the facility by the CRAC units, completing the airflow loop (see Figure 1.2).

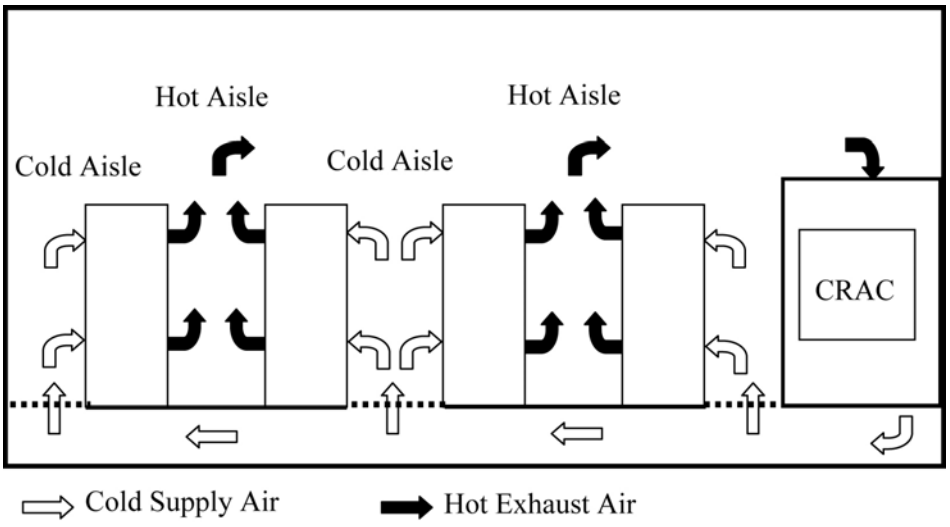


Figure 1.2. Standard raised floor plenum and room return cooling scheme

Thermal performance and efficiency of data centers could be improved by considering other possible supply and return schemes. To reduce the entrainment of hot air into the cold aisles, the exhaust could be collected above the hot aisle and ducted back to the CRAC units. Additionally, the cold supply air could be delivered to the racks from

above the cold aisle, which could also reduce the recirculation effects. Finally, both the supply and return could be overhead.

1.3. Systems-Level Electronics Thermal Management

Traditional electronics thermal management studies have focused on the heat transfer characteristics of individual packages and multiple devices in an enclosure such as an individual server. Systems-level electronics cooling is defined here as data processing systems comprised of multiple enclosures or modules, including data processing racks containing various servers and data center facilities containing numerous racks. Figure 1.3 illustrates the current representative volumetric heat generation rates at various length scales in electronics components and systems. It is the heat generation at the interconnect level that causes the temperature rise between the chip and the ambient and drives the need for a cooling system to maintain the chip operating at an acceptable temperature. In systems-level thermal management, the cooling scheme is often developed at the largest length scales, such as the data center CRAC unit. Thus, thermal modeling and characterization is inherently multiscale, as it is the heat generated at the smallest length scales that drives the need for a global cooling scheme at the largest length scales.

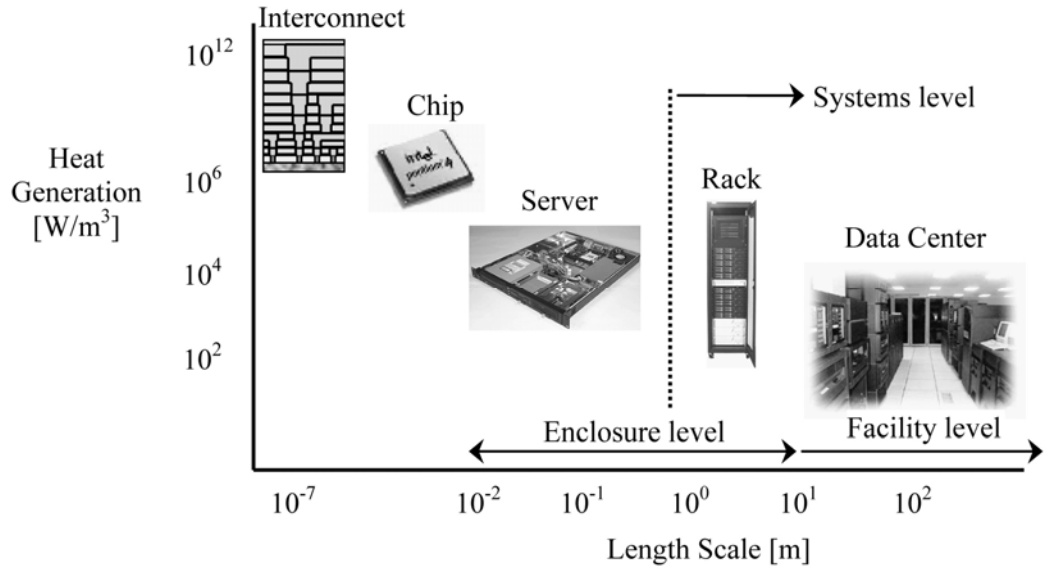


Figure 1.3. Systems-level electronics thermal management heat generation and length scale map

1.4. Objectives of Data Center Modeling and Characterization

Modeling of data centers aims to predict the dominant airflow and heat transfer characteristics so a reliable design methodology may be developed. An accurate modeling framework can be used to create more thermally efficient data center facility designs, eliminating local hot spots and using the cold supply air from the CRAC units in the most efficient way possible. Currently, overall energy balances on the facility are used as a design guideline. The expected net power dissipation of the equipment is computed and an appropriate number of CRAC units is selected to provide the cooling. Considering the rapid increase in power density and the high reliability requirements, designers often over-specify the amount of cooling required in a data center. The lifetime of a data center is approximately 30 years, whereas the lifetime of the data processing equipment inside the facility is only about 2 years, requiring several cycles of equipment replacement, each with higher performance and power. Besides the efficient design of new data centers, improved modeling and characterization also aims to address questions

that arise from such equipment upgrades in an existing data center. For instance, if a new server that has twice the power dissipation and airflow requirements is to be added to a rack, how will this affect neighboring servers? A robust modeling strategy can be used to predict if this new server will deprive neighboring servers of supply air and how the hot exhaust air will affect nearby severs. Model efficiency is a key concern as traditional numerical procedures for turbulent flows and heat transfer will produce excessively computationally intensive models, severely limiting their use as an analysis and design tool.

1.5. State of the Art and Future Trends in Data Center Thermal Management

With continually increasing power dissipation in high performance computing equipment, the limits of forced air cooling will be reached in the near future. Over the past several years, supplemental thermal solutions in the form or ceiling mounted air cooling units and additional rack level fans have been used to deal with high power density at the rack level. Currently, data center professional have moved away from adding more air cooling units in favor of liquid cooling solutions.

Liquid cooling has the ability to reduce the power consumption and noise levels in the data center although reliability is a major concern, especially if the cooling fluid is not dielectric. Commercially available CRAC units universally use chilled water or glycol as the heat transfer medium between the data center air to the environment. The piping infrastructure to supply the CRAC units with cooling liquid exists in the data center environment. Most liquid cooling schemes tap into this liquid loop to provide additional cooling at the rack level. Such systems commonly use air-to-liquid heat

exchangers mounted inside the rack to cool the hot exhaust to form a self-contained cooling loop. This approach reduces the distance the hot exhaust air must travel before reaching the heat exchanger, which minimizes the effects of recirculation. Moving the heat exchanger to the rack still relies on air as the heat transfer medium and the performance of such systems are usually limited by the air side heat transfer coefficient. To date, there is ongoing research and development of extending the liquid cooling path to the chip level for maximum efficiency, but no commercially available products exist.

Recently multiple vendors have manufactured liquid cooling solutions to be used at the rack exhaust. The objective is to cool the exhaust air to minimize the effect of a high power dissipation rack on surrounding racks. Supplemental liquid cooling is not used at the rack inlet because of the possibility of condensation near data processing equipment.

2. NUMERICAL MODELING OF DATA CENTERS

Due to the complex nature of the flow inside a data center, computational fluid dynamics and heat transfer (CFD/HT) are usually required to investigate the thermal performance of data centers. The problem is inherently multiscale, with the smallest scale phenomena driving the global temperature field. It is the heat dissipation at the chip level that causes the temperature rise across each electronic device within each server. A rack contains a number of servers, each with their power dissipation and airflow requirements. The chip may contain features as small as 10 nm, and the facility length scale is on the order of 100's of meters, implying a computational model resolving all the features in a data center would contain 10 decades of length scales. It is computationally impossible to resolve all these features, but a numerical model should attempt to resolve as many length scales as necessary for the type of predictions desired.

2.1. Flow Regimes and Scaling

The flow inside a facility typically falls into the turbulent mixed convection regime, although there can be large spatial variations in turbulence intensity and buoyancy effects. The flow through the data processing equipment inside a rack is mainly forced convection, but buoyancy effects may become important at the rack exhaust and as the hot air returns to the CRAC units. Reynolds numbers can be defined using different length scales and velocity scales and the range of both these scales in data center airflows makes quantifying airflow characteristics with relatively few dimensionless groups difficult. Data center thermal performance is typically considered

using a steady state approximation because a large factor in cooling and power equipment redundancy makes failures of such systems a rare event.

As stated earlier, electronics cooling applications span a wide range of length scales. With current computing power, numerical simulations of heat transfer and fluid flow are limited to approximately three decades of length scales. This has limited chip-level thermal analysis to resolving some larger features contained within the package (die and heat spreaders) but the smallest features are still inaccessible. Board-level simulations typically employ a simplified representation of the packages, treating them either as a block with uniform heat generation or as a constant heat flux area. Numerical computations of electronic enclosures, such as personal computers and servers, are able to resolve multiple boards and devices, but only the components with the largest power dissipation are usually considered. Modeling infrastructures of electronic equipment must bridge the large disparity in length scales between the board and the facility.

An Electronics Industry Association (EIA) standard rack measures 0.74 m wide and 0.94 m deep [1]. The vertical direction of a rack is standardized in racks units call "U", where 1 U = 4.45 cm (1.75 inches) and a typical rack can accommodate 40 U of equipment. Consider a rack containing 10 servers, each measuring 4 U high and dissipating 500 W. A flow rate of 0.497 kg/s is required to maintain a 10 °C bulk air temperature rise, resulting in a hydraulic diameter based Reynolds number for a single server of 5,730. The hydraulic diameter is calculated by assuming the server is 0.74 m wide by 0.18 m (4U) tall rectangular duct. The Rayleigh number, based on the same 10 °C temperature difference and a characteristic length of (duct cross-sectional area) / (duct perimeter) = 0.0787 m is 4.46×10^5. Using Ra / Re^2 as an estimate of the ratio of

buoyancy to inertia effects (which is valid because $Pr \approx 1$ for air) gives 0.0136, showing that buoyancy can be neglected based on such global consideration. As the rack power dissipation increases, the required flow rate will also increase to maintain constant component temperatures, causing the Reynolds number to increase, while the Rayleigh number remains the same, making the ratio of Ra / Re^2 decrease. These scaling arguments have also been experimentally justified by demonstrating point-wise measurements of rack inlet temperatures are linearly proportional to the CRAC exhaust temperature [6].

2.2. Review of Data Center Numerical Modeling

Data center facilities are critical to the data processing needs they serve. They are designed for continuous operation in order to provide a highly reliable environment for the data processing equipment performing essential tasks and often containing highly sensitive material. Access to these environments is limited and performing detailed experimental measurements in operational data centers is very difficult. Thus, most previous investigations in data center thermal management are computational in nature and rely on CFD/HT to predict airflow and heat transfer characteristics.

CFD/HT modeling of data centers was introduced in 2001 by Patel *et al.* [1]who performed computations on a model facility and by Schmidt *et al.* [7] who compared experimental measurements through raised floor data center perforated tiles with 2-dimensional computational models. The various numerical investigations can be classified into the following 4 main categories: 1) raised floor plenum airflow modeling to predict perforated tile flow rates, 2) thermal effects of rack layout and power

distribution and 3) investigation of alternative supply and return schemes and 4) development of thermal performance metrics. A review of data center literature is presented by Schmidt and Shaukatullah [8], which serves to provide a historical perspective considering the rapid growth in data center power density.

Schmidt *et al.* present a depth-averaged (2-dimensional) numerical model for a 0.284 m deep plenum [7] and for a 0.292 m deep plenum [9]. Their experimental validation shows fair overall agreement with select tile flow rates, with large individual prediction errors. In both papers, it is noted that a 0.1 m diameter pipe and a 0.025 m tall cable tray are located in the bottom of the plenum, although there is no discussion of how these obstructions are accounted for in the depth-averaged equations. Schmidt *et al.* [9] state that for plenum depths less than 0.3 m deep, depth-averaged modeling is adequate as a tradeoff between computational cost and accuracy. Radmehr *et al.* [10] experimentally investigated the leakage flow, or portion of total CRAC airflow that seeps through the seams of the raised floor tiles, in a 0.42 m deep facility. Distributing the leakage flow uniformly throughout the perforated tile seams and modeling chilled water supply lines, Radmehr *et al.* [10] were able to produce predictions with an overall accuracy of 90%. Van Gilder and Schmidt [11] present a parametric study of plenum airflow for various data center footprints, tile arrangements, tile porosity and plenum depth. A vacant plenum was assumed, which makes the perforated tile resistance much greater than any other flow resistance in the plenum and allows the resulting perforated tile flow rate to scale with the net CRAC flow rate.

Regarding data center rack layout, Patel *et al.* [1, 12] have focused on the aisle spacing and the corresponding cooling load on the CRAC units in conventional raised

floor plenum data centers. These efforts identify the problem of recirculation, or hot exhaust air being drawn into a cold aisle and through electronic equipment before returning to the CRAC units. The effects of recirculation are also documented by Schmidt *et al.* [9, 13-15]. Geometrical optimization of plenum depth, facility ceiling height and cold aisle spacing for a single set of CRAC flow rates and uniform rack flow and power dissipation was performed by Bhopte *et al.* [16]. The results showed that increasing the plenum depth or the facility ceiling height, and decreasing the cold aisle width produce an overall reduction in the average inlet temperature to all the racks in the facility. Schmidt and Iyengar [15] have also documented thermal performance variability in several operational data centers and attempted to provide some qualitative explanations for similarly performing racks in different data centers. A cluster of high-power dissipation racks and their effect on the remaining equipment of the facility were considered by Schmidt and Cruz [14]. The effect of varying power throughout a data center is an important factor to consider, as operational facilities contain a wide variety of equipment, each with their own power dissipation levels.

Each data center has a unique geometrical footprint and rack layout and a common basis is needed to compare the thermal performance of various cooling schemes. A unit cell architecture of a raised floor plenum data center was formulated by considering the asymptotic flow distribution in the cold aisles with increasing number of racks in a row [17]. The results indicated that for high flow rate racks, 4 rows of 7 racks adequately models the hot-aisle cold-aisle configuration and is representative of a 'long' row of racks.

One of the earliest works in data center CFD/HT modeling analyzed a global flow configuration other than that of the standard raised floor plenum with return to the CRAC unit through the room. The model of Patel *et al.* [1] examined a cooling strategy where the CRAC unit supply and return are located over top of the cold aisles. A novel concept of introducing cold supply air into the hot aisles to mitigate recirculation effects was proposed by Schmidt and Cruz [13] although computational results showed this modification could diminish the basic cooling scheme's efficiency. Supplying the cold air from above the racks was investigated in [18] where the overhead supply and facility height were investigated parametrically. Shrivastava *et al.* [19] parametrically vary all possible locations of the cold air supply and hot exhaust return for fixed room geometry, uniform rack power and fixed CRAC conditions.

In the previous analysis of data center airflows, researchers have either modeled the rack in a black-box fashion with prescribed flow rate and temperature rise, or as a box with fixed flow rate and uniform heat generation [1, 12, 14, 18]. This less computationally expensive approach is useful only if facility quantities such as CRAC unit performance and rack layout are of interest. However, this level of description cannot determine temperature variations within the rack caused by recirculation effects. To better characterize the performance variations through a data center, a procedure to model individual servers within each rack was developed in [20]. As computational limitations allow, the server sub-models can be increased in complexity to include discrete heating elements to mimic multiple components on a printed wiring board.

A formal modeling approach to data centers considering modeling of the CRAC units, perforated tiles and rack sub-models was presented in [21]. This investigation

considers the distribution of CRAC units through the unit cell architecture and the effect of rack orientation relative to the CRAC units. To develop a mechanistic understanding of convective processes in data centers, the global scheme was divided into the processes of 1) the CRAC exhaust to the cold aisle, 2) cold aisle distribution of supply air through the racks and 3) the hot exhaust air return to the CRAC units [22]. Numerical modeling of various supply and return schemes, coupled with various orientations between the racks and the CRAC units, identified the causes of recirculation and non-uniformity in thermal performance throughout the data center. The models presented in [1, 12, 14, 15, 18, 23] have used a variety of orientations between the CRAC units and racks. The parametric study presented in [21] is the first attempt to generally quantify these effects and following work using the same procedures was contributed in [19].

The performance of the assorted data center models are assessed in various ways by different authors and specific comparisons can be difficult. Sharma *et al.* [23] introduce dimensionless numbers to quantify the effects of recirculation. These dimensionless numbers are arrived at by considering the ratios (cold supply air enthalpy rise before it reaches the racks) / (enthalpy rise at the rack exhaust), and (heat extraction by the CRAC units) / (enthalpy rise at the rack exhaust). These definitions require the air temperature evaluation at arbitrary points near the rack inlet and exhaust. Sharma *et al.* [24] later computed these dimensionless performance metrics in an operational data center by taking a single air temperature measurement just below the top of each rack inlet and outlet. Norota *et al.* [25] used the statistical definition of the Mahalanobis generalized distance to describe the non-uniformity in rack thermal performance. Shah *et al.* [26-28] have proposed an exergy-based analysis method that divides the data center

into a series of subcomponents and then CRAC unit operation is optimized given information regarding the rack power dissipation.

Thermal performance metrics for systems level electronics cooling based on the concept of thermal resistance (power dissipation / temperature rise) were formulated and applied to data centers in [21]. The metrics consider the spatial uniformity of thermal performance to characterize poor designs causing local hot spots. Entropy generation minimization was also proposed as a data center thermal performance metric because poor thermal performance is often attributed to recirculation effects. Since the mixing of hot exhaust air with the supply air in the cold aisles generates entropy, cold aisle entropy generation minimization was employed as a metric. The results presented in [21] show that using entropy generation minimization and thermal resistance with spatial uniformity considerations predict the same design as being the best. The same thermal performance metrics were applied to a forced air-cooled rack to optimize the server layout [29] because the metrics were formulated generally for systems-level thermal management and should be applicable to a range of systems.

2.3. Limitations of Numerical Modeling

In the absence of detailed experimental data on data center airflow and heat transfer, CFD/HT provides a tool to estimate some of the dominant features of data center thermal management. The overall accuracy of models needs to be addressed due to many simplifications that are employed. RANS-based turbulence modeling procedures come into question for the strongly swirling flow in data centers. All the numerical models employed the standard k-ε turbulence model [30], except for [21, 22, 29, 30] where the

full Reynolds stress model [30] was used to assess the isotropic eddy viscosity assumption.

Given highly accurate modeling procedures, the CFD/HT models are still limited due to the large amount of time invested in constructing a mesh, demonstrating mesh and iteration convergence and finally post-processing the results. The models presented in [21, 22] exceed 1.5 million grid cells and complicated pressure-driven boundary conditions require many iterations for the model to converge. For a 2-equation turbulence model in 3 dimensions, the finite volume method produces 7 degrees of freedom (DOF) per grid cell (*P, u, v, w, k, ε,* and *T*) or 13 DOF with the addition of the 6 Reynolds stress components in the full Reynolds stress model. The total amount of time consumed by a single solution, including mesh development, grid convergence and post-processing was approximately 100 hours, with 80 hours dedicated computing time on a dual Pentium Xeon 2.8 GHz processor workstation with 4 Gb of memory. All of the DOF solved for are necessary to model the airflow and heat transfer, but the key quantities of interest such as rack temperature rise and perforated tile flow rates, are evaluated using only a small portion ($<10\%$) of the total model DOF.

3. REDUCED-ORDER MODELING OF TURBULENT FLOWS

Design and analysis of complex engineering systems involving turbulent convection often require careful numerical simulations using computational fluid dynamics and heat transfer (CFD/HT) or detailed experimental data obtained by full field techniques such as tomographic interferometry and particle image velocimetry, to accurately describe the fluid flow and heat transfer processes of the system [31]. Both methods of system characterization are time intensive, particularly so for carrying out parametric studies, and severely limit the range of design variables that can be explored, rendering optimization algorithms impractically slow. In early stages of design, it may be desirable to trade this effort for an experimentally-validated reduced model that captures the dominant physics, but is computationally efficient. Such models may be used in conjunction with optimization routines to quickly perform parameter sensitivity studies. A low-dimensional model of this type can also be integrated with multiscale computations to efficiently bridge a range of length scales without requiring a single simulation to resolve all length scales simultaneously.

3.1. Reduced-Order Model Taxonomy

Methods of reduced-order modeling can generally be divided into state space and distributed parameter system approaches. State space methods reduce a system to an input/output map and many tools are available to analyze interconnected state space components, see [32] for an application-based overview. This class of models is also synonymous with ‘lumped-parameter model’ and can utilize basic physical principles

such as mass and energy conservation as well as correlations to develop the model behavior. Figure 3.1 below schematically depicts a state space model of a server that returns the maximum chip temperature and mean air outlet temperature given the inlet temperature, chip power dissipation and flow rate. The maximum chip temperature can be evaluated using local Nusselt number relations $Nu_x = f(Re_x, Pr, Gr_x)$ and Newton's law of cooling, $Q = h_x A(T_x - T_\infty)$. The model is based on a fixed geometry and material properties. Resistor network type models are often used to model electronic packages and are categorized under state space methods as they typically return a junction temperature given the power dissipation, an external heat transfer and ambient conditions [33-37]. The fixed parameters are the chip geometry and materials, which strongly affect the manner in which the thermal resistance network is constructed.

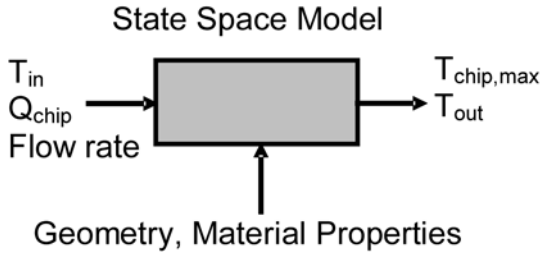


Figure 3.1. State space model of a server

Models of this form may provide an adequate level of accuracy, but the description of the underlying physical mechanisms is very incomplete. State space models cannot generally be examined to determine the root causes for specific behavior and their performance strongly depends on the modeling assumptions they are built upon.

Distributed parameter systems aim to approximate the physics over the entire domain, as opposed to returning a vector of desired outputs. This approach is desirable in

modeling convective flows, as the model is not limited to returning a prescribed quantity, such as a set of velocity, temperature or heat flux information, rather the complete velocity and temperature fields are available for further analysis of the transport processes involved. The fundamental principle of distributed parameter modeling is to find a suitable set of modes to project the governing equations onto, reducing the solution procedure to finding the appropriate weight coefficients that combine the modes into the desired approximate solution. Traditional modal expansions using Fourier series or orthogonal polynomials (Legendre, Chebyshev, Laguerre) form the basis for spectral methods. Complex boundary conditions can be problematic in spectral methods and the types of boundary conditions often dictate the functions employed in the expansion. For example, Fourier series are the natural choice for periodic domains while the properties of Chebyshev polynomials are often exploited for inhomogeneous boundary conditions. Higher order terms are required to resolve sharp gradients in the solution and many terms in the series are needed for accurate predictions.

The process of taking a model from a large number of DOF, either from detailed numerical simulations or full-field experimental measurements, to a model involving significantly fewer DOF is termed *model reduction*. A number of tools exist for reducing the number of internal states of large linear systems, such as those resulting from descretizing differential equations [32]. System identification is the process of identifying model structure or estimating unknown parameters through experimenting with an unknown system. Considering the black box type system depicted in Figure 3.1, system identification would aim to estimate $T_{chip,max}$ in the most accurate and efficient

manner by varying the system inputs. Figure 3.2 illustrates this taxonomy of efficient modeling procedures [32].

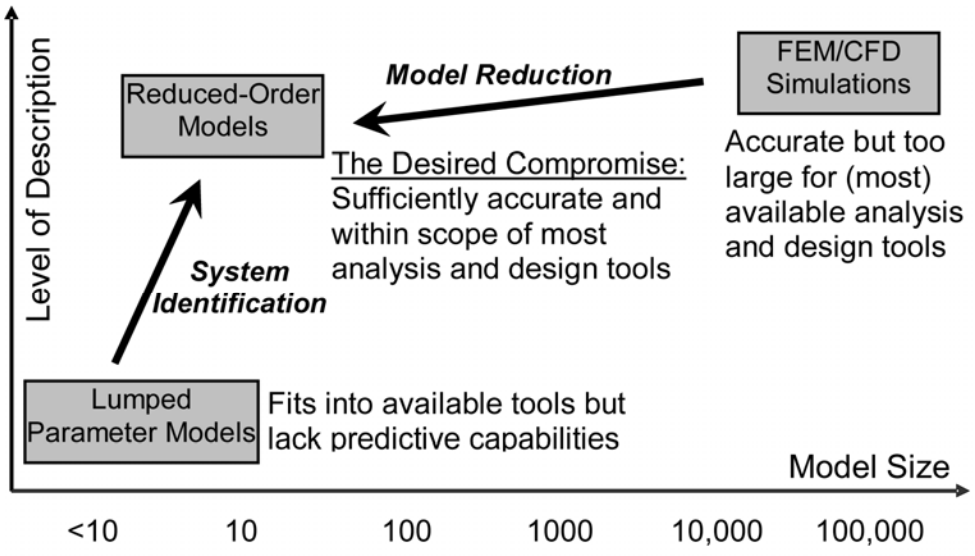


Figure 3.2. Model description and size comparison, from [32]

With objective of constructing a reduced-order modeling framework for systems-level electronics thermal management, a key model development would involve the construction of models for parametric systems. A *parametric system* is defined here as a system containing a source term or boundary condition that can be varied over a specified range in order to produce different system responses. For a convective flow, such parameters include a geometrical length, mass flow rate, boundary heat flux or temperature, or volumetric heat generation, to produce different flow patterns, and/or transport characteristics. This can be quantified with one or more relevant dimensionless groups, such as Reynolds or Strouhal number, non-dimensional heat generation rates, or a set of aspect ratios to define geometry. Changes in thermophysical property variations

are excluded from this definition because they may arise naturally even in the absence of the parametric variabilities defined here. In summary, the parametric nature may result from a prescribed boundary condition, for example inlet velocity or wall heat flux, interior condition, such as volumetric heat generation, or geometric parameter such as an aspect ratio.

3.2. The Proper Orthogonal Decomposition: Literature Review

The proper orthogonal decomposition (POD) is a stochastic tool used to assemble the model-specific optimal linear subspace from an ensemble of system observations. Owing to the stochastic nature of the subspace calculations, the POD is ideally suited for nonlinear phenomena and has been used extensively in low-dimensional modeling of turbulent flows, see [38] for a more complete description and review of its use in prototypical turbulent flows. A major shortcoming of the existing POD methodology to date is its inefficient treatment of a range of model parameters. Previous reduced-order flow and heat transfer modeling studies have investigated the dynamics of a prototypical system under a single Reynolds or Rayleigh number or limited range of variation. Laminar flows were investigated by Deane *et al.* [39] and *ad hoc* mode scaling showed mixed results (± 15% in period predictions) for approximating flows from 52 observations over a small range of Reynolds numbers. For laminar flows with various inlet profiles, both Park and Kim [40] and Ravindran [41, 42] suggest homogenizing the POD modes by subtracting a reference field that satisfies the governing equations.

Park and Kim [40] constructed a low-dimensional controller for flow over a backward facing step based on 1,000 observations of 2 inlet velocity profiles. Ravindran

homogenized 100 observations for developing a flow controller using blowing at Re = 200 [41, 42]. In the investigation of transitional behavior, Ma and Karniadakis used 40 modes to study the limit cycle of 3-dimensional flow around a cylinder at a critical Re = 188 [43] and at Re = 610 [31]. No special boundary treatment was required for either flow because periodic boundary conditions were employed.

In low-order modeling of heat transfer, Park and Cho partitioned the linear governing equation into homogeneous and inhomogeneous components to account for boundary conditions in order to model conduction [44] and temperature and species transport for a fixed velocity field [45], using 200 and 400 observations respectively. Sirovich [46-48] analyzed the dynamics of natural convection by working with homogeneous deviations from the mean flow, allowing the mean to take the fixed inhomogeneous boundary conditions and typically using around 200 observations. Park and Li [49] modeled natural convection with 30 sinusoidal boundary heat flux profiles for a total of 3,000 observations.

Recent developments in low-dimensional flow modeling have been made by Sirisup and Karniadakis [50] who have proposed using a penalty function Galerkin method to treat time varying boundary conditions. Geometrical scaling has also been investigated by Taylor and Glauser [51] who constructed a low-dimensional model of a variable angle diffuser at the expense of 30,720 observations. Uttakar *et al.* [52] used POD for reduced turbulent simulations of flows with moving boundaries; however they do not describe any reduced-order model development, only the accuracy and data compression associated with POD representation of the observations. Galletti *et al.* [53] modeled laminar flow over a confined square block by interpolating the modal weight

coefficients at different Reynolds number to correct the pressure drop across the duct from 160 observations. To summarize the work to date in developing POD-based models of flows over a range of varying Reynolds and Rayleigh numbers, inhomogeneous boundary conditions are either treated through expensive homogenization procedures or through a very large, and often impractical, number of system observations.

A key concern in the existing POD methodology is determining the minimum number of observations required to construct a POD subspace that faithfully represents the physics of the system. In dynamic systems, each simulation time step is available to be included in the ensemble. Experimentally-based POD models of turbulent flows also benefit from large data ensembles, as many repeated measurements are required to generate statistically significant turbulence data. For parametric reduced-order modeling of stationary turbulent flows, each observation is from an independent system snapshot. To the best of the author's knowledge, the only other published attempt of using POD-based modeling for stationary analysis of thermo-fluid systems was by Ly and Tran [54]; however their solution method was based on interpolating splines between weight coefficients to match a desired parameter value. This method would require higher order multi-dimensional interpolation to model a complex system with multiple parameters and also does not guarantee that the desired parameter level will be achieved.

There are two major deficiencies with the standard POD procedure as described above. In the context of dynamical systems, the Galerkin projection has been demonstrated to produce false limit cycles [55] and deemphasize important modal contributions under varying bifurcation parameters in parameter dependent flows [56, 57], ultimately leading to unphysical results. Secondly in previous reduced-order flow

modeling studies, homogeneous boundary conditions in the form of either closed [48, 58] or periodic domains [39, 43] are employed. Inhomogeneous boundary conditions have also been treated by subtracting reference velocity fields from each member of the ensemble to homogenize the boundary conditions before the POD modal basis is computed [42]. The reference field must satisfy the governing equations, implying that if the boundary conditions are to be altered a new reference field for each set of boundary conditions must be obtained.

3.3. Mathematical Formulation

The POD uses principal component analysis to decompose a large DOF system into a series of fundamental modes and an approximation to the governing equations is sought using the expansion theorem:

$$u(x,t) = \sum_{i=1}^{m} a_i(t)\varphi_i(x) \tag{3.1}$$

Solution methods based on (3.1) require the specification of a family of functions forming the modal basis $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_m\}$ that span the domain Ω. The basis functions usually satisfy homogeneous boundary conditions individually and inhomogeneous boundary conditions are treated by adding a source function:

$$\vec{u}(x,t) = \vec{u}_0(x,t) + \sum_{i=1}^{m} a_i(t)\vec{\varphi}_i(x) \tag{3.2}$$

Note in the most general case the source function, $\vec{u}_0$, may account for time dependent boundary conditions. In the context of fluid flow, the solution and modal basis will be considered vector-valued functions. In traditional POD analysis of turbulent flows, the

source function in (3.2) often assumes the form of the ensemble mean, $\vec{u}_0 = \langle \vec{u} \rangle$, which renders the POD modes akin to the Reynolds stresses.

The POD is a stochastic tool that computes the optimal linear basis for the modal decomposition in (3.2). Given an ensemble of system observations, $\{\vec{u}_k \in H \mid k = 1,2, ...,m\}$, $\vec{\Phi}$ can be computed by minimizing the projection error onto the ensemble, $\min\{\langle \| \vec{u}_k - P_k \vec{u}_k \| \rangle\}$ where P_k is the orthogonal projector, $\|\cdot\|$ is the induced norm on the Hilbert space H and $\langle \cdot \rangle$ denotes the ensemble average. This is equivalent to maximizing the energy (in the sense of the induced norm) of the projection of the observations onto the basis functions [38, 59]. Standard variational calculus methods to extremize the functional $\left\langle \left| (\vec{u}, \vec{\varphi}) \right|^2 \right\rangle - \lambda(\|\vec{\varphi}\|^2 - 1)$, where the $(\|\vec{\varphi}\|^2 - 1)$ term is included to produce a normalized basis and λ is a Lagrange multiplier, show the POD basis vectors are the eigenvectors of the vector-valued eigenvalue problem:

$$\int_\Omega \vec{R}(x, x')\vec{\varphi}(x')dx' = \lambda_k \vec{\varphi}(x') \tag{3.3}$$

where $\vec{R}(x, x') \equiv \left\langle \vec{u}(x) \otimes \vec{u}^*(x') \right\rangle$ is the cross –correlation function and $(\cdot)^*$ denotes the complex conjugate. The associated eigenvalue with each empirical basis vector $\lambda_k = \left\langle \left| (\vec{u}_k, \vec{\Phi}) \right|^2 \right\rangle$ is a measure of the average energy (in the sense of the induced norm) of the projection of the ensemble onto $\vec{\Phi}$ because $\vec{R}$ is a self-adjoint linear operator $R: H \to H$.

For discrete data, $\vec{R}(x,x')$ is computed by taking *m* observations of the system containing *n* DOF and assembling them into the observation matrix

$\vec{U} = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_m\} \in \mathbb{R}^{n\,x\,m}$. The empirical basis functions (referred to as 'POD modes') are then the eigenvectors of the covariance matrix $(1/m)\vec{U}\vec{U}^T \in R^{n\,x\,n}$, which may be computationally infeasible for large DOF problems since eigenvalue algorithms can typically only handle matrices on the order of 10^5, rendering the POD impractical for large data ensembles. This problem can be circumvented by realizing $span\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_m\} = span\{\vec{\varphi}_1, \vec{\varphi}_2, ..., \vec{\varphi}_m\}$ and the POD basis can be expressed as a linear combination of the linearly independent observations, known as the method of snapshots, see [60], *viz*:

$$\vec{\varphi}(x) = \sum_{i=1}^{m} b_i \vec{u}_i(x) \tag{3.4}$$

The weight coefficients in (3.4) are eigenvectors of the solution to $\vec{R}'\vec{\varphi} = \lambda\vec{\varphi}$ where $\vec{R}' = (1/m)\vec{U}^T\vec{U} \in R^{m\,x\,m}$, allowing the number of observations and not the number of system DOF to dictate the size of the basis function computation. Using the method of snapshots to assemble the basis functions as admixtures of the system observations implies that $\vec{\Phi}$ intrinsically contains any linearly invariant properties of $\vec{u}_k$ from the fact that $\vec{\Phi}$ has been computed only through linear operations on $\vec{U}$. Thus, the POD modes individually satisfy the incompressibility condition, $\nabla \cdot \vec{\varphi}_k = 0 \;\; \forall \; k$ and homogeneous boundary conditions, $\vec{\varphi}(\partial\Omega) = 0$ where $\vec{u}(\partial\Omega) = 0$. For the incompressible flow considered here, H is the space of smooth, vector-valued, solenoidal functions on Ω equipped with the inner product $(\vec{u}, \vec{v}) \equiv \int_\Omega \vec{u} \cdot \vec{v} dx$ and $(\vec{u}, \vec{v}) \equiv \sum_i u_i v_i$ for discrete vectors.

The eigenvalues of the kernel in (3.4) are positive semi-definite because the kernel is positive semi-definite, however, eigenvectors belonging to the null space $\lambda = 0$ are

ignored because they contain no information about the system in the sense $\lambda_k = \left\langle \left| (\vec{u}_k, \vec{\Phi}) \right|^2 \right\rangle = 0$. This leads to an incomplete $L^2(\Omega)$ space but poses no problem because the subspace only needs to describe the physics contained within the observation set $\vec{u}_k$ and including all admissible basis functions as would directly oppose the reduced-order modeling framework. In the computation of the POD modes, a numerical cut-off criterion on the order of the machine precision is used to eliminate eigenvectors associated with $\lambda < O(10^{-12})$, as these modes are laden with numerical error [56].

From an implementation standpoint, the POD mode may be computed by assembling the observation matrix $\vec{U} = \{\vec{u}_1, \vec{u}_2, ..., \vec{u}_m\} \in \mathbb{R}^{n \times m}$ and then decomposing $\vec{U}$ using the singular value decomposition (SVD). Given a matrix $A \in \mathbb{R}^{n \times m}$, the SVD produces the decomposition $A = U\Sigma V^T$. $U \in \mathbb{R}^{n \times n}$ is a matrix whose columns form the left singular vectors of A, $V \in \mathbb{R}^{m \times m}$ is a matrix whose columns form the right singular vectors of A and Σ is a pseudo-diagonal matrix of the singular values. It can be shown that $A^T A = V\Sigma^2 V^T$ and that the eigenvalues in (3.4) are equal to $\lambda_k = \sqrt{\Sigma_{kk}}$. The method of snapshots can be implemented as $B = SVD(\vec{U}^T \vec{U})$ with $B \in \mathbb{R}^{m \times m}$ and then assembling $\vec{\Phi} = \vec{U}B$. The energy captured by each POD mode is then computed as $E_k = \frac{\lambda_k}{\sum_{j=1}^{m} \lambda_j}$ and the total energy resolved using the first p modes $E_{1 \to p} = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{j=1}^{m} \lambda_j}$.

3.3.1 The Galerkin Projection

The standard method of evaluating the weight coefficients in (3.2) is to project the governing equations onto the modal subspace, known as the Galerkin method or Galerkin

projection. Consider the 3-dimensional incompressible flow governed by the Navier-Stokes without external forcing and inhomogeneous boundary conditions over a portion of the domain boundary:

$$\nabla \cdot \vec{u} = 0 \tag{3.5a}$$

$$\vec{u} \cdot \nabla \vec{u} - \nu \nabla^2 \vec{u} + \frac{1}{\rho} \nabla P = 0 \tag{3.5b}$$

$$\vec{u} \mid_{\partial\Omega} = \vec{u}_b \tag{3.5c}$$

where ρ is the fluid density and ν is the kinematic viscosity. The Galerkin projection of the modal basis onto the governing equations is computed by taking the L^2 inner product between the two, recalling the integral definition, $\langle \vec{u}, \vec{v} \rangle \equiv \int_\Omega \vec{u} \cdot \vec{v} dx$:

$$\int_\Omega \vec{\varphi}_i \cdot (\vec{u} \cdot \nabla \vec{u} - \nabla^2 \vec{u}) + \frac{1}{\rho} \nabla P) dx = 0 \tag{3.6}$$

This procedure orthogonalizes the residual to each basis function of the modal subspace. Substituting the modal expansion in (3.2) onto (3.6) yields the *m*-dimensional system of equations (3.7a):

$$C_{ijk} a_j a_k - D_{ij} a_j + A_{ijk} a_j a_k + S_i + B_i = 0 \tag{3.7a}$$

$$C_{ijk} \equiv \int_\Omega \vec{\varphi}_i \cdot (\vec{\varphi}_j \cdot \nabla \vec{\varphi}_k) dx \tag{3.7b}$$

$$D_{ij} \equiv \nu \int_\Omega \vec{\varphi}_i \cdot \Delta \vec{\varphi}_k dx \tag{3.7c}$$

$$A_{ijk} \equiv \int_\Omega \vec{\varphi}_i (\vec{u}_o \cdot \nabla \vec{\varphi}_k + \cdot \vec{\varphi}_k \cdot \nabla \vec{u}_o) dx \tag{3.7d}$$

$$S_i \equiv \int_\Omega \vec{\varphi}_i (\vec{u}_o \cdot \nabla \vec{u}_o - \nu \Delta \vec{u}_o) dx \tag{3.7e}$$

$$B_i \equiv \frac{1}{\rho} \int_\Omega \vec{\varphi}_i \cdot \nabla P dx \tag{3.7f}$$

where summation over repeated indices is implied. The C_{ijk} term results from the convective operator ($\vec{u}\cdot\nabla\vec{u}$), D_{ij} from the diffusive term ($\nu\nabla^2\vec{u}$), A_{ijk} is the cross term between the source function and the POD modes, S_i comes the source term only and B_i is the projection of the POD modes onto the pressure term, ($\frac{1}{\rho}\nabla P$).

The pressure term B_i can be integrated by parts to obtain $B_i = \int_{\partial\Omega} P\vec{\varphi}_i \cdot \hat{n}dx$ with the aid of the divergence theorem and the incompressibility property of $\vec{\varphi}$. In (3.7b), the pressure term is a Lagrange multiplier to satisfy the continuity constraint (3.5a) and therefore should be unnecessary over the domain because $\nabla\cdot\vec{\varphi}_k = 0 \ \forall\, k$ by construction. The pressure on the boundary has physical significance because it provides the driving force for the flow. The main obstacle for inhomogeneous boundary conditions is the treatment of the boundary pressure and specifically coupling the pressure to the velocity a least squares manner as □ EMBED Equation.3 □□□ where □ EMBED Equation.3 □□□ isons, B_i = 0 because $\vec{\varphi}\cdot\hat{n}|_{\partial\Omega} = 0$ and for periodic boundary conditions, $\int_{\partial\Omega} P\vec{\varphi}_i \cdot \hat{n}dx = 0$ if there is no mean pressure gradient, such that the boundary term in (3.7a) is eliminated. For flows with inhomogeneous boundary conditions, some authors [40-42, 49] homogenize the boundary conditions by a reference field ($\vec{u}^r$) such that $(\vec{u}_k - \vec{u}^r)_{\partial\Omega} = 0$ before computing the POD modes to eliminate the need for boundary pressure-velocity coupling. As noted earlier, $\vec{u}^r$ must satisfy (3.5a-c) and needs to be redefined for each new inhomogeneous boundary condition, which could possible require another full numerical computation for each new set of boundary conditions.

The variational formulation of the POD basis functions produces the Sobolev space $\vec{\varphi} \in H^1$ with first derivatives belonging to $L^2(\Omega)$ and the use of homogeneous boundary conditions further restricts the approximation to the Soboloev space of homogenous functions $\vec{\varphi} \in H_0^1$. However, this need not be the case in Galerkin methods, as the tau method was first introduced by Lanczos in 1938 and presented in a more modern context by Gottlieb and Orszag [61]. The tau method is a spectral method that does not require the modes to satisfy the boundary conditions individually, but the sum of the weighted modes together must satisfy the boundary conditions. As noted by Rempfer [55], the Galerkin system in (3.7a-e) could potentially be solved subject to constraints in an attempt satisfy the inhomogeneous boundary conditions although these constraints limit the solution space to smaller subspace of $\vec{\Phi}$ where only admissible combinations of the POD modes satisfying the inhomogeneous boundary conditions reside.

3.3.2. Analysis of the RANS Equations

In modeling and computation of industrial turbulent flows, the Reynolds-averaged Navier-Stokes equations (RANS) are commonly used because the more detailed descriptions of large eddy simulation and direct numerical simulation (DNS) are computationally impractical for complex geometries and boundary conditions. The RANS equations model the effect of turbulence on the mean flow as a spatially dependent effective viscosity and the steady momentum equation in the absence of body forces is:

$$\vec{u} \cdot \nabla \vec{u} - \nabla \cdot (\nu_{eff} \nabla \vec{u}) + \frac{1}{\rho} \nabla P = 0 \tag{3.8}$$

The velocity, $\vec{u}$, appearing in (3.8) represents the ensemble-averaged velocity field in the Reynolds decomposition $\vec{U} = \vec{u} + \vec{u}'$ where $\vec{U}$ is the total velocity and $\vec{u}'$ time varying deviation from the ensemble mean. Equation (3.5a) serves as the continuity equation in the RANS system for the incompressible flows considered here. There are many different eddy viscosity closure schemes, with the standard *k*-ε model and full Reynolds Stress model being discussed earlier [30].

Examination of the turbulent kinetic energy (k) and turbulent energy dissipation rate (ε) transport equations as well as the Reynolds stress transport equations shows that these equations model the effect of turbulence as a function of local velocity gradients only, i.e. $\nu_{eff} = f(\nabla\vec{u})$. Equation (3.8) can be viewed as a laminar flow model of a fluid with a strain-rate dependent viscosity. This modeling eliminates all of the small-scale turbulent dynamics, making it reasonable to expect that the required number of observations and modes to construct a low-dimensional model is similar to that of a laminar flow rather than a highly detailed set of DNS data, which often incorporate thousands of observations.

In the coupled phenomena of natural convection, Sirovich and Tarman [48] demonstrate that the velocity and temperature field can be decomposed separately in lieu of working with an extended state vector of the form $\vec{y} = \{\vec{u}, \nu_{eff}\}$. Assuming the a similar decomposition for the velocity and eddy viscosity fields, $\vec{u} = \sum_{i}^{m} a_i \vec{\varphi}_i$ and $\nu_{eff} = \sum_{j}^{m} b_j \psi_j$, the Galerkin projection of the POD modes onto the RANS momentum equation yields:

$$\int_{\Omega} \vec{\varphi}_i \cdot (\vec{u} \cdot \nabla \vec{u} - \nabla \cdot (\nu_{eff} \nabla \vec{u}) + \frac{1}{\rho} \nabla P) dx = 0 \tag{3.9a}$$

$$C_{ijk} a_j a_k - D_{ijk} b_j a_k + A_{ijk} a_j a_k + S_i = 0 \tag{3.9b}$$

$$D_{ijk} \equiv \int_{\Omega} \vec{\varphi}_i \cdot (\psi_j \cdot \nabla \vec{\varphi}_k) dx \tag{3.9c}$$

$$S_i \equiv \int_{\Omega} \vec{\varphi}_i \cdot (\vec{u}_o \cdot \nabla \vec{u}_o - \nabla \cdot (\psi_o \nabla \vec{u}_o)) dx \tag{3.9d}$$

where the C_{ijk}, $A_{ijk,}$ and B_i terms are the same as in (3.7a-f) and the boundary pressure term can be ignored at this point without loss of generality.

In the analysis of (3.5a-c), the governing equations are second order spatially. As was previously ignored, the POD modes, and therefore the resulting approximations, belong to the space H^1. In laminar flow approximations, this poses no problems because the smoothing action of the Laplacian implies that solutions starting in H^1 will rapidly enter H^2, [38]. This is a potential problem for the RANS equations, as no such smoothing is guaranteed because the behavior of ν_{eff} is unknown *a priori*. The diffusive term D_{ijk} could possibly be integrated by parts to achieve a term that contains only first derivatives, however this operation will then require $\nu_{eff}|_{\partial\Omega}$ to be specified. Specifying the correct boundary conditions for the turbulence transport equations on inhomogeneous regions such as inflows and outflows can be difficult for any flow computation and must often be justified for complex flows.

Another large issue in using the Galerkin projection for the RANS equations appears when the turbulence transport equations are used to couple the eddy viscosity modes to the momentum equation. The standard *k*-ε model will be used here, but the arguments equally apply to other turbulence models. Additional modes describing *k* and

ε would need to be generated from the flow database, increasing the computational cost. The fundamental problem can be seen in the definition of the effective viscosity:

$$\nu_{eff} = \nu + c_\mu \frac{k^2}{\varepsilon} \tag{3.10}$$

Assuming modal decompositions for k and ε of the form $k = \sum_i g_i \zeta_i$ and $\varepsilon = \sum_i h_i \eta_i$, substituting these terms into (3.10) would result in the term $\sum_i g_i \zeta_i \sum_j g_j \zeta_j (\sum_k h_k \eta_k)^{-1}$

But η_k is a vector and its inverse is undefined. Terms of the form ε/k also arise in the ε transport equation. To avoid special treatment of the inverse terms that arise from the need to close the effective viscosity weight coefficients, refer back to comments regarding the nature of the RANS equations as a model for the laminar flow of a fluid with a strain-rate dependent viscosity.

3.4. Flux Matching Procedure

To improve the computational efficiency of POD analysis by avoiding the Galerkin projection and special treatment of inhomogeneous boundary conditions, a flux matching procedure is introduced. Since the POD modes are themselves solutions to the governing equations (3.5a-c and 3.8), new solutions can be generated by using inhomogeneous modes to satisfy the boundary conditions, loosely based on the tau method. The structure of the POD modes lets them resolve the flow field over the remainder of the domain. Since the POD basis is the optimal linear subspace for a set of $\vec{u}_k$, the approximate analysis is treated as a linearized problem rather than using the Galerkin projection to construct a possibly unstable nonlinear model. The implicit

assumption is the modal contribution can be uniquely determined by satisfying the boundary conditions.

Ensuring that only the boundary conditions are satisfied and ignoring the residual over the rest of the domain is a familiar concept. Using information from only a small subset of the entire domain to estimate the velocity field is the basis for flow control because it is impractical to distribute sensors through the entire flow field. One example used linear stochastic estimation to correlate pressure measurements at the wall and POD modal coefficients in order to develop a low-dimensional flow controller [51].

In the analysis of complex flows, the exact velocity profile on the boundary is often unknown. However, the design objective is often based on integral conditions, such as the appropriate mass flux through a portion of the boundary, prompting the introduction of the flux function:

$$F(\vec{u},\beta) = \int_{\Gamma} \rho\beta\vec{u}\cdot\hat{n}dx \quad , \quad \Gamma \subseteq \Omega \tag{3.11}$$

This function generally returns a vector because Γ may be a finite set of discontinuous surfaces. Depending on the transport problem at hand, the flux function could describe the flow of any scalar, such as mass ($\beta = 1$), energy ($\beta = E$), or species concentration ($\beta = c_i$), but mass flux is considered here to demonstrate the methodology and is denoted F_m. Note momentum is not considered a viable term in the flux function of the linearized analysis because is depends on the profile. Knowing the mass flux only is not sufficient to evaluate the momentum flux because $\langle u \rangle^2 \neq \langle u^2 \rangle$. In the absence of any knowledge about the specific velocity profiles on Γ, the mass flux is an ideal choice to construct velocity field approximations, as will be detailed in further sections.

To construct a new approximate solution, the fluxes can be expressed as the vector of goals $G \in \mathbb{R}^q$ corresponding to the desired mass flux through the set of control surfaces $\Gamma = \{\Gamma_1, \Gamma_2, ... \Gamma_q\}$ which define the desired flow field $\vec{u}^*$ such that $G = F(\vec{u}^*)$. The solution procedure is then to find the set of weight coefficients that minimizes the error on the set Γ:

$$min\{||G' - \sum_{i=1}^{p} a_i F_m(\vec{\varphi}_i)||\} \quad where \;\; G' = G - F_m(\langle \vec{u} \rangle) \tag{3.12}$$

The modal summation is carried to $p \leq m$ because the optimal approximate solution may require less than the total number of modes available. The weight coefficients can generally be computed as $a_i = \langle \vec{u}^*, \vec{\varphi}_i \rangle$, which in the flux matching procedure can be solved as $a = F_m^+(\vec{\varphi}_i) G'(\vec{u}^*)$, where $F(\vec{\varphi}) \in \mathbb{R}^{m \, x \, q}$ is the matrix obtained by operating (3.11) on the q control surfaces of the m POD modes and $()^+$ is a suitably defined generalized matrix inverse.

The solution procedure is carried out as a series expansion with the ordered POD modes forming the expansion sequence and terms are successively added to the series until the boundary conditions are satisfied. Algorithmically, this can be expressed as:

$$\Delta G_{m,i} = G_{m,i-1} - F_m(\vec{u}_{i-1}) \tag{3.13a}$$

$$a_i = F_m^+(\vec{\varphi}_i) \Delta G_{m,i} \tag{3.13b}$$

$$\vec{u}_i = \vec{u}_0 + \sum_{j=1}^{i} a_j \vec{\varphi}_j \tag{3.13c}$$

where $F^+ \equiv (F^T F)^{-1} F^T$ is the Moore-Penrose matrix pseudo-inverse producing the least squares approximation [62]. To initiate the calculation, $\Delta G_{m,1} = G_m - F_m(\vec{u}_0)$ and the

first modal weight coefficient is computed as $a_1 = F_m^+(\vec{\varphi}_1)\Delta G_{m,1}$. The process is repeated until the desired set of mass fluxes is satisfied. The solution process is akin to a perturbation expansion where the source function, $\vec{u}_0$ in (3.2), acts as the leading order solution and each modal contribution serves as the next order correction. The difference from traditional perturbation methods is that the successive corrections occur in state space to satisfy the desired mass fluxes, without considering the remainder of the domain. This is a highly desirable property as the exact velocity profile on the boundary of a complex flow may be unknown. Since POD modes are solutions to the governing equations (to within a multiplicative constant), they contain physically correct velocity profiles and satisfy the divergence free condition. There is no need to account for the pressure term in (3.8) as it can be viewed as a Lagrange multiplier to enforce the divergence-free condition of (3.5a), which is already satisfied by $\vec{\Phi}$ because $\nabla \cdot \vec{\varphi}_k = 0 \ \forall \ k$. This also implies that the approximations in (3.12) satisfies an overall mass balance on the domain because if $\nabla \cdot \vec{\varphi}_k = 0 \ \forall \ k$, then $\int_\Omega \nabla \cdot \vec{\varphi}_k dx = \int_{\partial\Omega} \vec{\varphi}_k \cdot \hat{n} dx = 0 \ \forall \ k$ by the divergence theorem. This formulation is a significant improvement over the work of Galletti *et al.* [53] who used POD to model flow over a block in a channel. The pressure drop through the channel varied with Reynolds number and a linear correction of the form $-(\nabla p, \vec{\varphi}_k)$ was incorporated to correct a Galerkin system of the form in (3.7a-f).

For parameter-dependent flows (as defined in §3.1), taking the system reference point to be the ensemble mean may not produce the best results over a range of parameters because different modes may become more or less important in

approximating the flow field as parameter values change. Using (3.12) with the ensemble mean as the source function for the approximation would only produce accurate results for solutions near the ensemble mean, $G \approx F_m(\langle \vec{u} \rangle)$, but the solution method must be robust in that it constructs accurate solutions over the entire range of parameters present in POD subspace.

To alleviate the poor approximations for solutions being far from the system reference point in parameter dependent flows, Graham and Kevrekidis [57] proposed taking averages over arc lengths in phase space and Christensen *et al.* [56] proposed pre-weighting certain modes to increase their contribution on the superposition. Following [56], observations could be repeatedly added into the ensemble to shift the system reference point in order to satisfy the inhomogeneous boundary conditions with the ensemble mean acting as the source function:

$$F_m(\langle \vec{u}^* \rangle) \to G \tag{3.14}$$

The immediate problem is that weighting is not unique as there are multiple sets of w_i that solve $F(\langle w_1\vec{u}_1 + w_2\vec{u}_2 + ... + w_m\vec{u}_m \rangle) = G$ and additional information about which modes to weight is required. A more significant problem lies in the fact that as the weighting factor on any one member increases, the modal spectrum asymptotically approaches $\lambda = \{1,0,0,...\}$, collapsing the POD subspace to a point near that single observation. Higher order modes are prematurely excluded in (3.12) because $G' \approx 0$ making the weight coefficients in $a_i \cong 0$ for $i > 1$, resulting in the loss of information from all other modes.

Approximations can be improved in light of (3.14), but all the features present in $\vec{\Phi}$ must be retained and accessible to (3.12) in order to develop a robust solution

methodology. To accomplish this, decompose the POD subspace into orthogonal complement subspaces:

$$\vec{\Phi} = \vec{\varphi}^{\perp} + \vec{\varphi}' \ , \ where \ \vec{\varphi}^{\perp} \in R^{n x s} \ \ and \ \ \vec{\varphi}' \in R^{n x m-s} \tag{3.15}$$

This decomposition is referred to as the PODc from here on. In a parametric flow, different modes should become more or less important under various parameter values [57]. The source term should then be a function of the mass flux goals, $\vec{u}_0(G_m)$, which can be accomplished by choosing members of the ensemble as the source function and constructing the POD subspace as an orthogonal complement. This method is superior to the standard mean-centered POD method (referred to as simply the POD from here on), where the source function is taken to be the ensemble mean because solutions 'far' from the mean tend to incur larger errors and this distance from the mean has previously been used as an error measure [56]. Also note, using the mean-centered POD modes in the above flux matching procedure to compute the modal weight coefficients may not satisfy the boundary conditions to the required accuracy for all required approximate solutions within the parameter will be satisfied.

The orthogonal complement $\vec{\varphi}^{\perp}$ is chosen to best satisfy the inhomogeneous boundary conditions ('forcing' modes) and $\vec{\varphi}'$ represents the flow features over the rest of the POD domain ('responsive' modes). Letting $\vec{u}^{\perp}$ denote the observations used to construct $\vec{\varphi}^{\perp}$, the modal expansion and minimization problem are modified to:

$$\vec{u}^{*} = \langle \vec{u}^{\perp} \rangle + \sum_{i}^{m} a_i \vec{\varphi}_i \tag{3.16}$$

$$min\{|| G' - \sum_{i=1}^{p} a_i F(\vec{\varphi}_i) ||\} \quad \text{where} \ \ G' = G - F(\langle \vec{u}^{\perp} \rangle) \tag{3.17}$$

The observations in $\vec{u}^{\perp}$ are be selected by the algorithm:

$$\min\{\left\|F(\vec{u}^{*}) - F(\vec{u}_k)\right\|_1\} \ , \ k = 1, 2, ..., m \tag{3.18}$$

This method is based on determining observations that are geometrically 'close' to $\vec{u}^{*}$ in the parameter space of *F*. Once the set $\vec{u}^{\perp} \subseteq \vec{u}$ is selected, the POD is performed on the orthogonal complement mean-centered observations $\vec{u}^{\perp} = \vec{u}^{\perp} - \langle \vec{u}^{\perp} \rangle$. The remaining observations $\vec{u} \not\subset \vec{u}^{\perp}$ are orthogonalized to $\vec{u}^{\perp}$ and the POD is performed without mean-centering. The full POD subspace is assembled as in (3.15) and reordered based on descending magnitudes of the combined eigenvalue spectrum.

Thus, the approximation procedure consists of selecting the 'closest' members of the ensemble to the desired solution to serve as the source function and the information about the flow physics contained in the remaining observations is converted into an orthogonal series expansion about the source function to make higher order corrections to the approximate solution. Essentially, the orthogonal complement subspace converts the POD basis into a purely spatial function to a parameterization of approximate solution, *viz*:

$$\vec{\varphi}(x) \Rightarrow \vec{\varphi}(x, G) \tag{3.19}$$

The size of the orthogonal complement subspace (*s)* generally depends on the density of observations. If enough modes are present to accurately represent the model behavior, *s = 1* will usually be sufficient. Increasing *s* to *2* will result in marginal approximation improvements and *s > 2* will generally degrade the approximation. The purpose of the orthogonal complement is to shift the source function in the POD subspace and using *s = 1* causes the source function to assume the form of the nearest observation.

3.5. Application to Laminar Flow

The methodology is demonstrated on two complex flow geometries, one laminar flow and one turbulent flow. The laminar flow situation is used because the proposed flux matching procedure can be directly evaluated against the Galerkin method. The turbulent flow situation is used to illustrate how the flux matching procedure can be used to develop low-dimensional models when there are modeling uncertainties, or even in situations where the governing equations are unknown. At this point, it is reiterated that just as with the standard POD, the method is applicable to both experimental and numerical data. Numerically generated data are used here for convenience and the numerical solutions are considered as 'exact' solutions, see [63] for a discussion on the effect inherent errors associated numerical and experimental data collection has on POD analysis.

Currently, the POD modeling methodology consists of 2 components; the POD mode calculation and the determination of weight coefficients, which are generally coupled through the boundary conditions. In most previous POD analyses, the components are decoupled through homogenous or periodic boundary conditions. The proposed reduced-order modeling methodology presented here significantly improves both of these components and develops a robust framework that satisfies parametric boundary conditions by coupling the two solution components. The first enhancement to the POD procedure is using orthogonal complement subspaces to satisfy the boundary conditions, denoted the PODc. The second development concerns the computation of the weight coefficients in (3.2). The flux matching procedure (FMP) was developed on physical arguments and its superior computational efficiency to the Galerkin projection

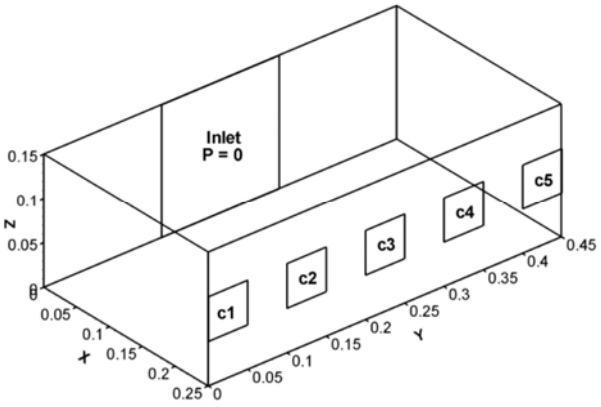
for steady flows will be demonstrated. Accordingly, there exist the following 4 combinations of possible solution procedures: the standard mean-centered POD with Gakerlin projection (denoted POD-GP), the orthogonal complement POD with Galerkin projection (PODc-GP), the POD with the flux matching procedure (POD-FMP) and combining the orthogonal complement POD with the flux matching procedure (PODc-FMP).

3.5.1 Laminar Flow Model Problem

Consider the flow in a manifold with a single inlet that distributes air at standard density to 5 outlet ports. The domain measures $5L$ x $9L$ x $3L$ with L = 0.05 m. The inlet measures $3L$ x $3L$ and the pressure constraint $P = 0$ was applied. The outlet ports measured L x L and were able to vary mass flow rates from 3.06 x 10^{-5} to 1.68 x 10^{-3} kg/s for a corresponding Reynolds number range of ($\mathrm{Re} = \dot{m} / \mu L$) 34 to 1888. These mass flow rates are based on an area-averaged velocity between 0.01 and 0.55 m/*s*. The final numerical model consisted of 135,000 grid cells and the 3 velocity components plus pressure solved for at each grid cell gives a total of 540,000 DOF. The mesh was shown to be convergent to less than 2% in terms of each velocity component and was demonstrated to be independent of iteration convergence criteria. The governing equations were solved using second order upwinding, the SIMPLEC pressure-velocity coupling algorithm and the PRESTO pressure interpolation scheme for rectilinear staggered grids [64].

Figure 3.3 schematically depicts the model geometry and defines the m = 10 randomly generated observations used to generate the POD modes. The observations are listed in terms of mass flux through the c_i outlet control surfaces according to (3.11)

because the objective of the flux matching procedure is to satisfy the net mass flux and let the structure of the POD modes determine the correct velocity profiles required by continuity.



Observation Database [kg/s]*1000

k	c1	c2	c3	c4	c5
1	0.8881	1.5313	1.4700	1.5619	1.1025
2	1.0106	1.1025	0.1838	0.2756	0.7350
3	1.0106	0.4900	1.4394	0.7044	0.0306
4	1.0719	1.0106	0.3675	0.1531	1.6844
5	0.4594	0.3675	0.0613	1.1025	0.9800
6	0.1531	1.1638	0.7963	0.0919	0.6738
7	0.4288	1.6844	1.1331	1.5925	1.1638
8	1.4394	0.7963	1.7763	0.1531	0.9494
9	0.9800	0.5206	0.6738	0.0919	0.9800
10	1.5313	0.2450	0.3063	0.1225	1.5006

Figure 3.3. Laminar flow model geometry and observation database

The number of system observations was arbitrarily chosen to be 10 as no rigorous theories exist for the minimum number of observations for fixed parameter flows or the distribution of observations for parameter-dependent flows needed *a priori* to construct the optimal data ensemble. This is an inherent limit in POD analysis and this investigation does not aim to extend those limits, rather it provides a computationally efficient methodology to treat a predefined range of inhomogeneous boundary conditions. To validate that enough system observations were made to construct a meaningful POD subspace, an additional 5 randomly generated observations were added to the data and the eigenvalue spectrum was recomputed. The first 9 eigenvalues of the mean-centered spectrum computed from the additional 5 observations did not change significantly, suggesting that the original 10 observations were adequate to construct the POD subspace.

The ensemble mean and the first 3 mode shapes generated by the basic mean-centered POD procedure are illustrated on the vertical mid-plane of the domain (z = 0.075 m) in Figure 3.4. The mode shapes demonstrate the decreasing energy of each POD mode and how lower order modes resolve dominant flow features while the higher order modes represent more localized effects. Also note that operating the mass flux function defined in (3.11) on the ensemble average produces a nearly uniform mass flow rate for each outlet port of 8.48 x 10^{-4} kg/s for a corresponding area-averaged velocity of 0.28 m/s. Thus, $F_m(\langle\vec{u}\rangle)$ produces a system reference point of uniform mass flow rates equal to the average value the parameter range spans, further indicating that a sufficient number of system observations have been made.

3.5.2. Laminar Flow Results

A test case of $G(\vec{u}^*) = \{10.41, 8.269, 7.044, 0.3063, 6.125\} \times 10^{-4}$ *kg/s* was randomly chosen and many other test cases were analyzed to ensure that the presented results are representative. Figure 3.5a-b plots the eigenvalue spectrum for both the POD and PODc procedure and the error associated with the satisfying the boundary conditions described by G and L^2 – norm of the error of the approximate velocity field. The error in matching the boundary conditions is defined relatively as $G_{err} = || F(\vec{u}^*) - G ||_2 \; / \; || G ||_2$ and likewise for the error over the domain, $u_{err} = || \vec{u}^* - \vec{u}_{exact} ||_2 \; / \; || \vec{u}_{exact} ||_2$.

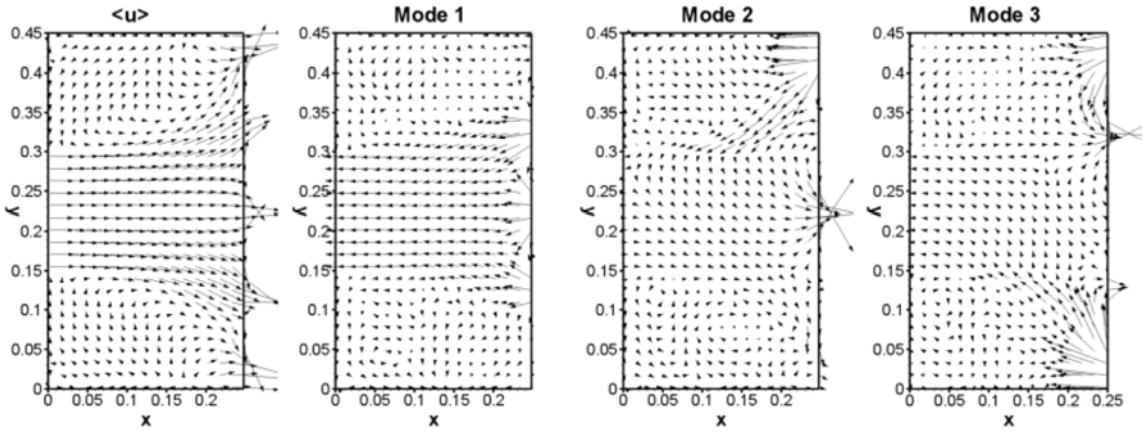


Figure 3.4. Laminar flow ensemble mean and the first three POD mode shapes, (n. b. the mode shapes have been interpolated to a grid 3 times less dense then the computational grid for illustrative purposes)

Figure 3.5a shows that the PODc procedure produces a steeper spectrum, indicating that fewer PODc modes are required to obtain a similar order of approximation accuracy as the POD modes. The PODc spectrum was computed using $s = 1$ and $s = 2$ observations to construct $\vec{\varphi}^{\perp}$, which were found to be observations $k = 9$ and $k = 2$, through the aid of (3.18).

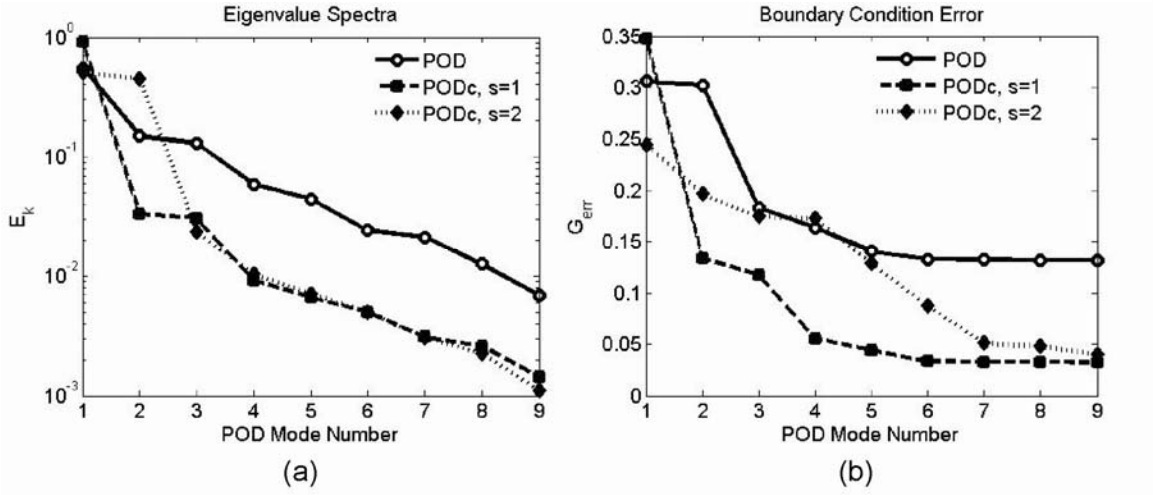


Figure 3.5. Approximate solution results, (a) modal eigenvalue spectra, (b) boundary condition error

Note the mean-centering procedure reduces the rank of the covariance matrix by one, therefore only m-1 modes have nonzero λ and are available for the reconstruction. In the PODc procedure, the mean-centering reduces the number of modes in $\vec{\varphi}^{\perp}$ by one, therefore, *s* must be greater than or equal to 2 produce 'forcing' and 'responsive' modes. Using $s = 1$ in the PODc causes the system reference point (source function) to correspond to a single observation and in the limit $s \to m$, the mean-centered POD procedure is recovered. Using $s = 1$ produces slightly more accurate approximations than $s = 2$ and increasing $s > 2$ does not significantly improve the approximation and increases the computation time, thus s = 1 is the best dimension for $\vec{\varphi}^{\perp}$.

The POD generates fixed modes independent of G. Figure 3.5a demonstrates there may exist a G that cannot be satisfactorily using the mean-centered POD. This is remedied by the PODc procedure, which essentially extends $F(\vec{\varphi})$ to $F(\vec{\varphi},G)$. For all test cases considered, , the boundary conditions can be satisfied to $O(10^{-2})$.

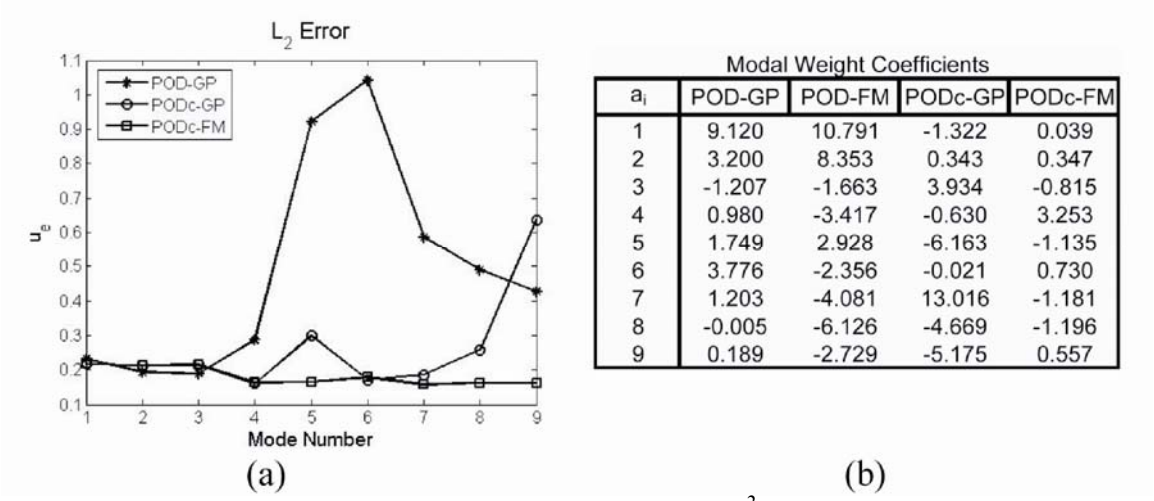
The solution method for the Galerkin system of equations deserves some comments. In previous POD fluid flow analysis, the low-dimensional models were created to investigate flow dynamics or to develop a control scheme. The resulting Galerkin system is a m-dimensional system of ODEs in time for the evolution of the modal coefficients, which are then numerically integrated. A similar approach could be taken for parametric steady flows where (3.7a-e) are integrated until a steady solution is obtained. The boundary pressure term in (3.7f) is set to zero because of the specialized treatment of inhomogeneous boundary conditions presented here. In the scope of reduced-order modeling and fast approximations, (3.7a-e) are solved directly because the

numerical integration would be computationally inefficient. The solution is posed as the linearly-constrained minimization problem:

$$\min_{a}\{C_{ijk}a_j a_k - D_{ij}a_j + A_{ijk}a_j a_k + S_i\} \text{ subject to } Fa - G' = 0 \qquad (3.20)$$

which is solved using standard sequential quadratic programming methods. The minimization problem is treated as a single objective optimization based on $||a||_\infty$. The minimum is defined as the location in the *m*-dimensional vector space of weight coefficients that when substituted into (3.7a) returns the minimum of $||a||_\infty$. The solution to $a = F(\vec{\varphi})^+ G'$ is used as the initial guess, and the convergence to a minimum from that point was verified by using other initial guesses. The termination criteria were also varied to ensure the solution was independent of the convergence tolerance.

Figure 3.6a plots u_{err} for the POD-GP, PODc-GP and PODc-FMP solution procedures while the results from the POD-FMP procedure are not shown because the boundary conditions are never satisfied. The results show using the Galerkin projection with the standard mean-centered POD produces large errors over the domain. The L^2 error norm (u_{err}) error is greatly reduced with the PODc modes and the flux matching procedure results in a decaying error while the Galerkin projection produces an error that increases with the addition of modes. Figure 3.6b lists the modal weight coefficients and it can be seen that the Galerkin projection and flux matching procedures produces a similar set of coefficients using the POD modes.



Modal Weight Coefficients

a_i	POD-GP	POD-FM	PODc-GP	PODc-FM
1	9.120	10.791	-1.322	0.039
2	3.200	8.353	0.343	0.347
3	-1.207	-1.663	3.934	-0.815
4	0.980	-3.417	-0.630	3.253
5	1.749	2.928	-6.163	-1.135
6	3.776	-2.356	-0.021	0.730
7	1.203	-4.081	13.016	-1.181
8	-0.005	-6.126	-4.669	-1.196
9	0.189	-2.729	-5.175	0.557

Figure 3.6. Laminar flow approximate solution, (a) L^2 error norm and (b) modal coefficients

Each result presented in Figure 3.6a for the Galerkin solution to the weight coefficient problem is based on a different system because C_{ijk}, D_{ij}, A_{ijk} and S_i change with m. Equation (3.20) cannot be solved for the full m modes and then only use the first $p < m$ to approximate the solution because the boundary conditions are only satisfied using the full number of modes from which the Galerkin system was computed. This may impose significant computational effort when deciding the number of modes to retain in the reduced-order model, as the numerical evaluation of C_{ijk}, D_{ij}, A_{ijk} and S_i is time consuming and needs to be performed for each p. The mode shapes in the PODc procedure differ from those in the POD, but the Galerkin projection and flux matching solution procedure produces similar modal weight coefficients for a given set of modes. These weight coefficients are listed in Figure 3.6b. The solution to the Galerkin system requires $p \geq q$ because there is no feasible solution to (3.20) if there are more constraints than the dimension of system.

Figure 3.7a illustrates contours of the absolute error superimposed on the vertical mid-plane of the approximate solutions for various numbers of modes computed from the PODc-FMP. The absolute error is plotted because the relative error becomes unrepresentatively large where the true velocity field goes to zero. The results show that the reduced-order model captures the dominant flow physics very well and the largest errors are incurred in regions of large velocity gradients near the outflow ports of the domain. Figure 3.7b plots isosurfaces of the absolute velocity error over the entire domain.

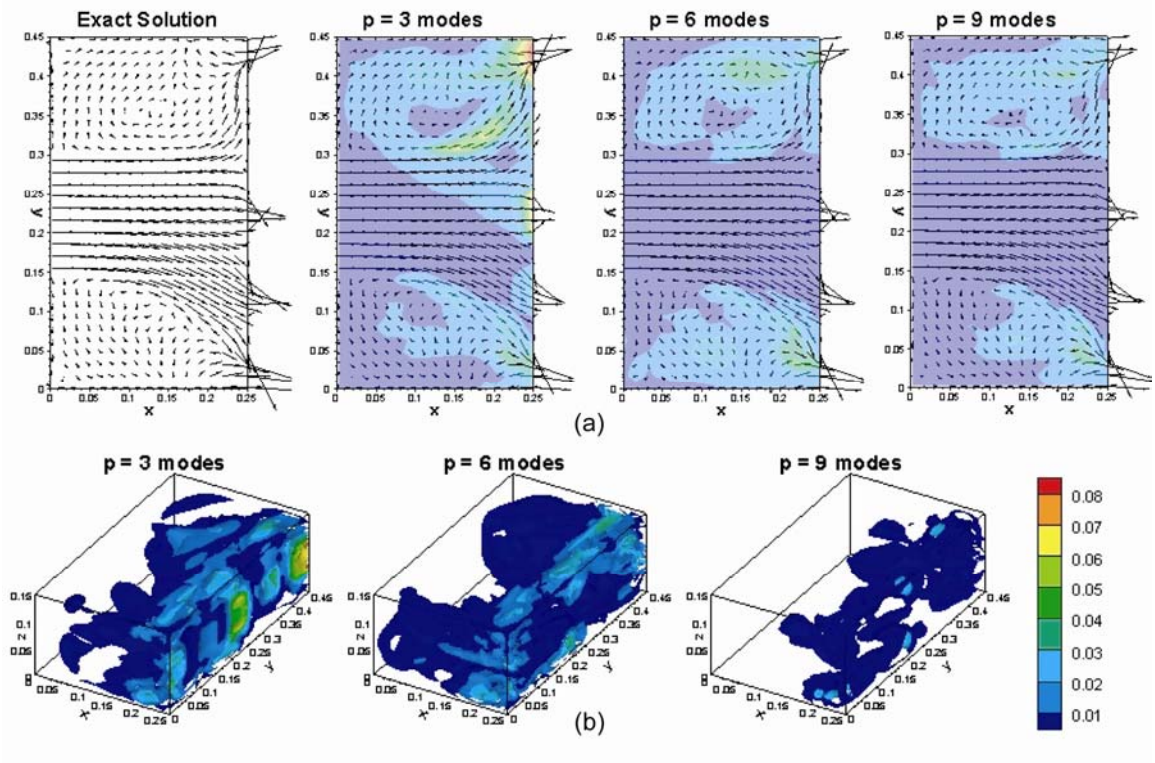


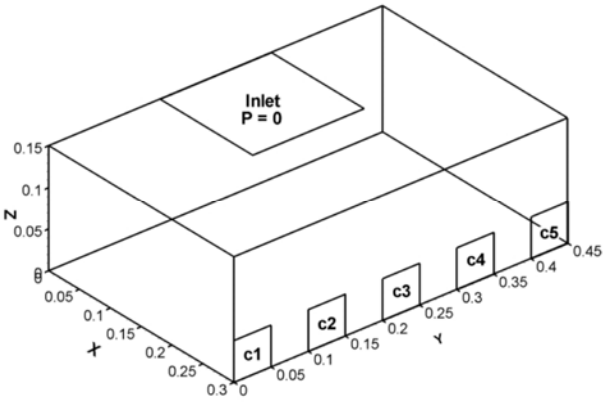
Figure 3.7. Error reduction with increasing system dimension (a) on the vertical mid-plane and (b) over the entire domain

3.6. Application to Turbulent Flow

As a turbulent flow example, consider a similar 3-dimensional turbulent flow of air at standard density and temperature in a single inlet manifold with five equal sized outlet ports. The outlet ports are specified to have mass fluxes ranging from 0.0092 – 0.025 kg/s, for corresponding hydraulic diameter based Reynolds numbers ranging between 10,269 and 27,383. The domain is $6L$ x $9L$ x $3L$ with the outlet ports measuring L x L and the inlet measuring $3L$ x $3L$ with L = 0.05 m. The full numerical model contained 162,000 grid cells with 10 DOF at each grid cell (3 velocity components, pressure and 6 Reynolds stress components) for a total 1,620,000 flow DOF. The mesh was shown to be convergent to less than 1% in terms of each velocity component and was demonstrated to be independent of iteration convergence criteria.

3.6.1. Turbulent Flow Model Problem

Figure 3.8 schematically depicts the model geometry and defines the m = 10 observations used to generate the POD modes. The first two observations were chosen to have the uniform minimum and maximum velocity of the parameter range and the remaining observations were generated randomly based on the corresponding area-averaged velocity rounded to the nearest integer.



Observation Database [kg/s]

k	c1	c2	c3	c4	c5
1	0.0092	0.0092	0.0092	0.0092	0.0092
2	0.0245	0.0245	0.0245	0.0245	0.0245
3	0.0153	0.0153	0.0123	0.0153	0.0153
4	0.0184	0.0092	0.0092	0.0153	0.0245
5	0.0245	0.0123	0.0092	0.0245	0.0123
6	0.0184	0.0245	0.0184	0.0245	0.0092
7	0.0092	0.0214	0.0153	0.0245	0.0214
8	0.0214	0.0153	0.0092	0.0092	0.0123
9	0.0153	0.0245	0.0153	0.0214	0.0153
10	0.0153	0.0153	0.0153	0.0245	0.0092

Figure 3.8. Turbulent flow model geometry and observation database

3.6.2. Turbulent Flow Results

A number of test cases were randomly chosen and the results shown here are for the set $G(\vec{u}^*) = \{1.74, 1.08, 1.16, 1.64, 1.58\} \cdot 10^{-2}$ kg/s , which is typical of other test cases. Figure 3.9a plots the eigenvalue spectra of the standard POD and PODc procedures and Figure 3.9b plots the boundary mass flux error. Figure 3.10 illustrates the PODc mode shapes on vertical mid plane (z = 0.075 m). The $\vec{u}^{\perp}$ subspace is assembled using $s = 2$ observations which were determined to be k = 3 and k = 4 from (3.18). Increasing $s > 2$ does not increase the accuracy of the approximate solutions and slightly increases the computational time, and again s =1 is optimal.

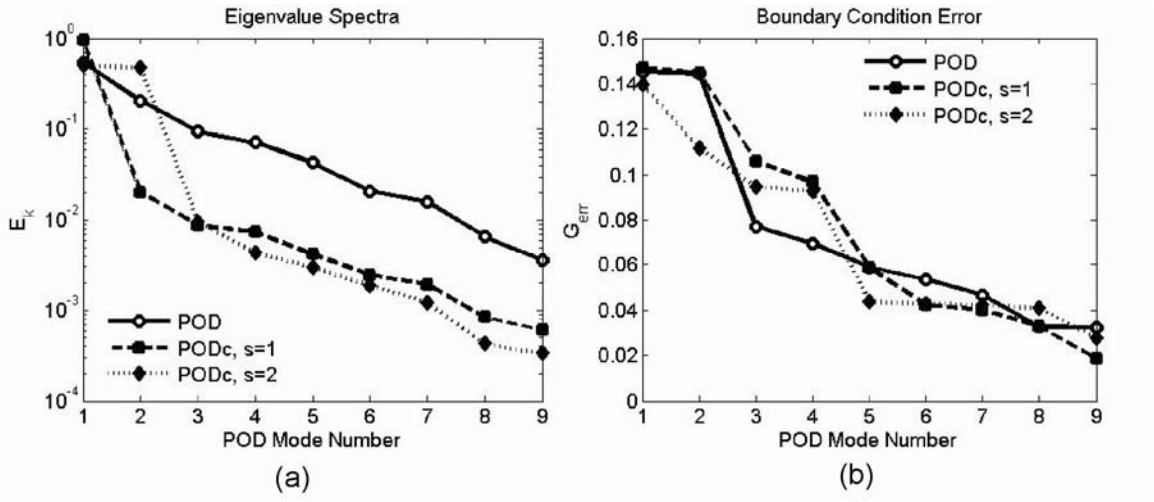


Figure 3.9. Approximate solution results, (a) modal eigenvalue spectra, (b) boundary condition error

The PODc produces two modes with relatively large eigenvalues, corresponding to the dominant driving and responsive modes while the remainder of the PODc spectrum exhibits a sharper decay relative to the POD spectrum, indicating an improvement in accuracy for approximations employing the same number of POD modes. Figure 3.10 illustrates that the dominant forcing mode contains sharper gradients and more secondary effects, but better reproduces the boundary mass fluxes. Since the mode shape is a solution to (3.8) up to a multiplicative constant, the resolved secondary effects are a direct cause of the boundary conditions.

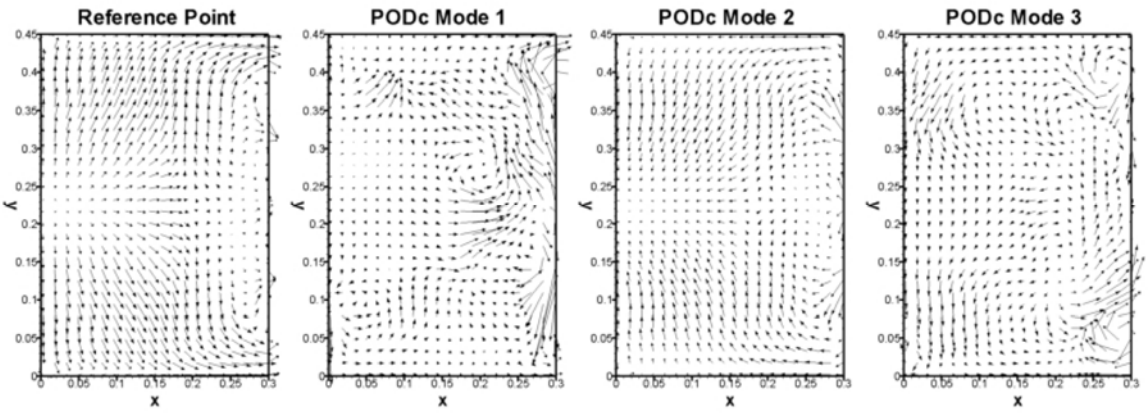


Fig 3.10. Turbulent flow system reference point and the first three POD mode shapes from the PODc

The error norm results in Figure 3.11a show that the standard POD-GP method produces poor results with increasing error as additional modes are added. The data show the FMP generally produces better approximations than the Galerkin projection and the orthogonal complement subspace PODc is a significant improvement over the mean-centered POD for treating parametric boundary conditions. The flux matching procedure reduces the L^2 error norm proportionally with the goal residual, indicating convergence. The PODc-FM outperforms the POD-FMP due to the segregation of the POD subspace into driving and responsive modes. The data in Figure 3.11a indicate the first $m \leq q$ modes will not reduce the error over the domain because the boundary conditions are not satisfied, but for $m > q$, the error is rapidly reduced, although no guarantee exists that the error will tend to zero as more modes are added to ensemble and used in the approximate solution.

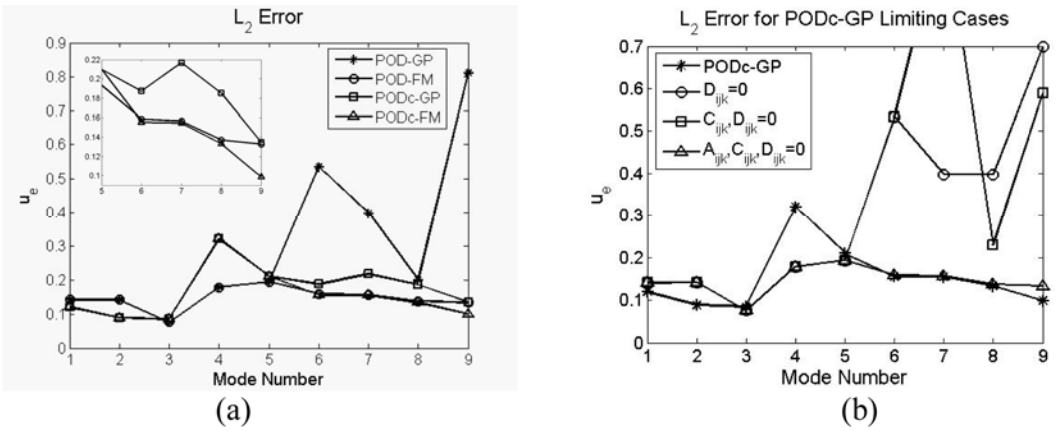


Figure 3.11. Turbulent flow approximate solution results (a) L_2 error and (b) Galerkin projection limiting cases for constant viscosity assumption.

Recall the RANS equations appear as a model for the laminar flow of a strain rate dependent viscosity fluid and that the results in Figure 3.11a demonstrate that using a constant viscosity in the RANS Galerkin projection produces a reasonably accurate approximation considering the level of simplification involved. Thus, it would be natural to consider some limiting cases of the Galerkin projection of the RANS equations, (3.9a-c). Since the flow is boundary driven and the source function S_i is used to satisfy the boundary conditions, one could argue that setting $C_{ijk} = D_{ijk} = A_{ijk} = 0$ would result in a good approximation to the full PODc-GP solution. Other limiting cases would be dropping only the diffusive term (D_{ijk}) since it is dominated by the convective term in high Re turbulent flows and to drop the convective and diffusive terms such that only the source term and the cross term (A_{ijk}) remain. The results presented in Figure 3.11b show that indeed the source function dominates the solution to Galerkin system, further justifying the use of the flux matching procedure, where only the boundary conditions are satisfied, over solving the full Galerkin system.

To further illustrate the error field of the approximate solution, Figure 3.12a plots the exact $\vec{u}^*$ solution and approximate solutions using increasing number of modes with superimposed u_e contours on the vertical mid-plane. Figure 3.12b plots isosurfaces of u_e over the entire domain. Figures 3.12a-b demonstrate that approximation errors occur in regions of large velocity gradients because the POD procedure orders the modes in descending energy and the small scale flow features are lost in the finite truncation of the modal expansion.

Determining the number of observations to generate the POD modes is a major concern of all POD analyses. An additional 5 randomly generated observations were added to ensemble and u_e showed approximately the same rate of convergence as in Figure 3.11 for $p > 5$ modes, implying the flux matching procedure converges as $p \to \infty$. In dynamic simulations each time step serves as an observation, however in steady parameter dependent flows, each observation is independent and may take significantly more time to generate. Thus, it is desirable to compute as few observations as necessary and it is noted that the 5 additional observations do not significantly change the eigenvalue spectrum, suggesting an adequate representation of the POD subspace.

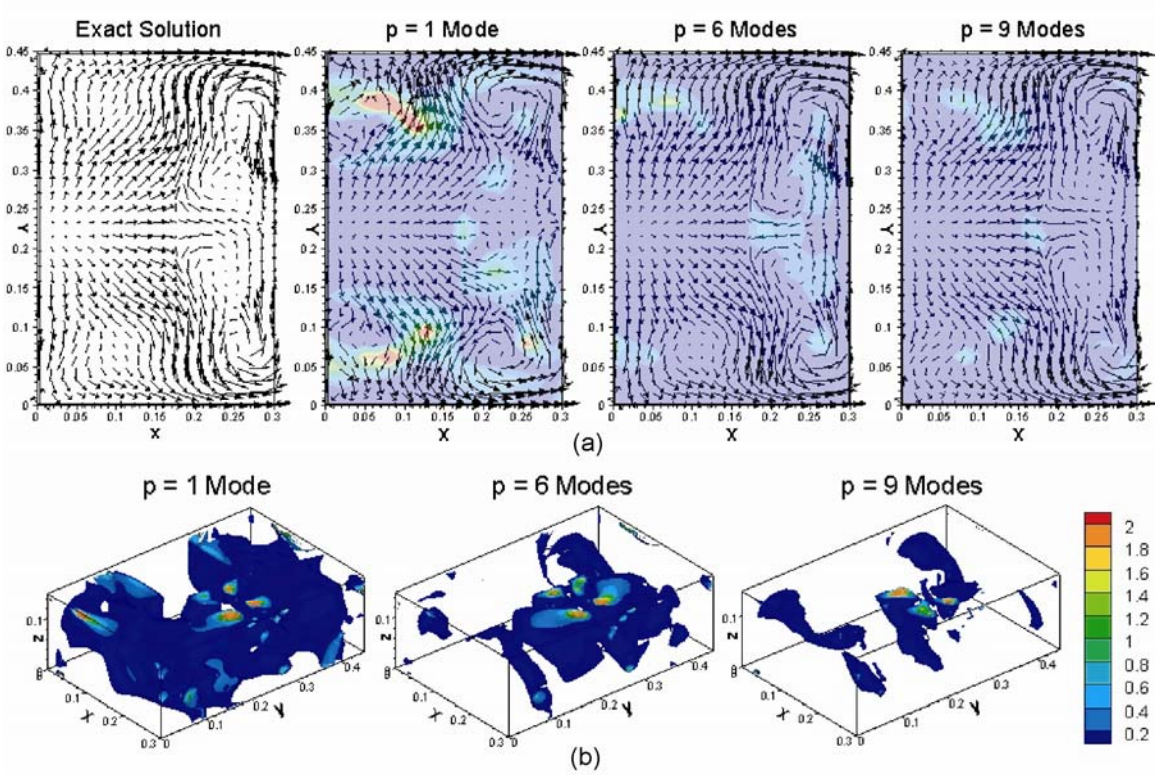


Figure 3.12. Error reduction with increasing system dimension (a) on the vertical mid-plane and (b) over the entire domain

3.7. Application to Air-Cooled Electronics Rack

To illustrate the PODc-FMP methodology for a representative problem encountered in systems-level electronics cooling, consider a 2-dimensional representation of an air-cooled data processing cabinet containing 10 servers. Cold supply air is delivered to the rack through a 0.39 m cutout in the bottom and is drawn into each individual server to maintain a safe operating environment for the data processing equipment. Each server contains an induced draft fan model to produce the necessary flow and lumped resistance at the inlet to account for the pressure drop across the server. Two 0.30 m tall by 0.50 m long blocks in each server are given a constant heat flux to mimic the power dissipation of high performance central processing units (CPUs). The rack dimensions are based on commercially available units and all walls are modeled as

adiabatic. Figure 3.13 below describes the rack and server geometry and schematically depicts the airflow patterns.

In the single parameter case, the server fan model is fixed to produce a nominal 0.180 kg/s (310 CFM) flow rate and each CPU dissipates 20 W for a total rack power of 400 W. The inlet velocity (V_{in}) is varied between 0.0 and 2.0 m/s and maintained at a constant 288 K. For the multi-parameter case, the inlet velocity is fixed at 0.5 m/s and the following 3 different types of servers are used: a low power with nominal 0.08 kg/s flow rate and 20 W per CPU, a medium power with nominal 0.13 kg/s flow rate with 30 W per CPU and a high power with 0.17 kg/s flow rate and 40 W dissipated per CPU. The rack is assumed to contain 3 low-powered, 4 medium-powered and 3 high-powered servers to limit the design space. Both of these cases illustrate real world type design problems. The single parameter objective is to find the optimal inlet flow rate to ensure reliable operation of all the servers in the rack. The multi-parameter case is an example of finding the thermally optimal arrangement of servers to minimize hot exhaust air from servers polluting the inlet supply of others [29].

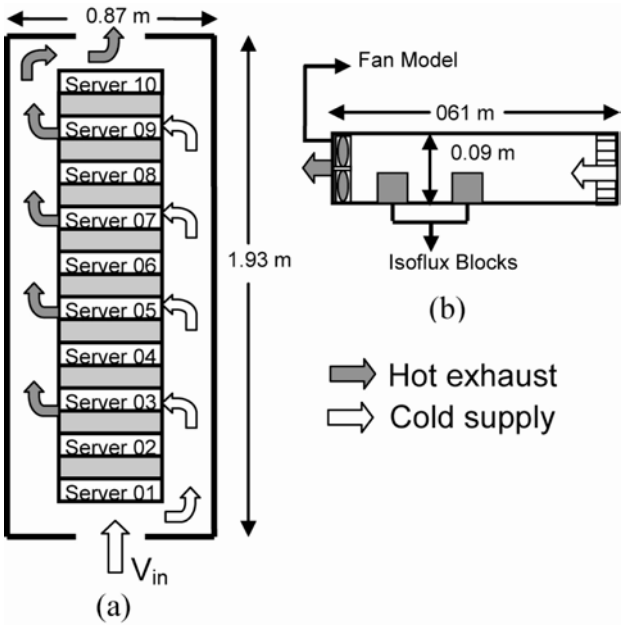


Figure 3.13. (a) Rack and (b) server geometry and airflow patterns

The full CFD model neglects buoyancy effects and solves the steady incompressible RANS momentum and energy equations with no body forces, using second order upwinding and SIMPLEC pressure-velocity coupling with PRESTO pressure interpolation [64]. The final converged model contained 20263 grid cells for a total of 121578 DOF. Note that the sum of the server flow rates is always greater than the rack net flow rate for all cases considered in the single and multi-parameter studies. This flow rate imbalance requires a large degree of recirculation, or a server's hot exhaust being drawn through another server before exiting the rack, and a correspondingly complex flow field.

3.7.1. Single Parameter RANS POD

The techniques to create reduced-order models of RANS based CFD/HT computations for facilitating design studies such as optimization are presented first for the single parameter flow case in order to clarify the development of the methodology. The results will then be extended to the multi-parameter case, which represents a pressure-driven flow. The observations for the single parameter case where created by varying the inlet velocity between 0.0 and 2.0 m/s on 0.25 m/s increments, $V^{obs} = \{0.0, 0.25, 0.50, 0.75, 1.0, 1.25, 1.50, 1.75, 2.0\}$ m/s, for a total of 9 observations,. Based on inlet hydraulic diameter and the minimum nonzero and maximum values of V^o, the Reynolds number of the observations ranges from 8,829 to 70,629. The normalized eigenvalue spectrum (λ) corresponding to the mean-centered POD modes is shown in Figure 3.14 below. Note velocity magnitude contours are shown to better illustrate the overall flow patterns.

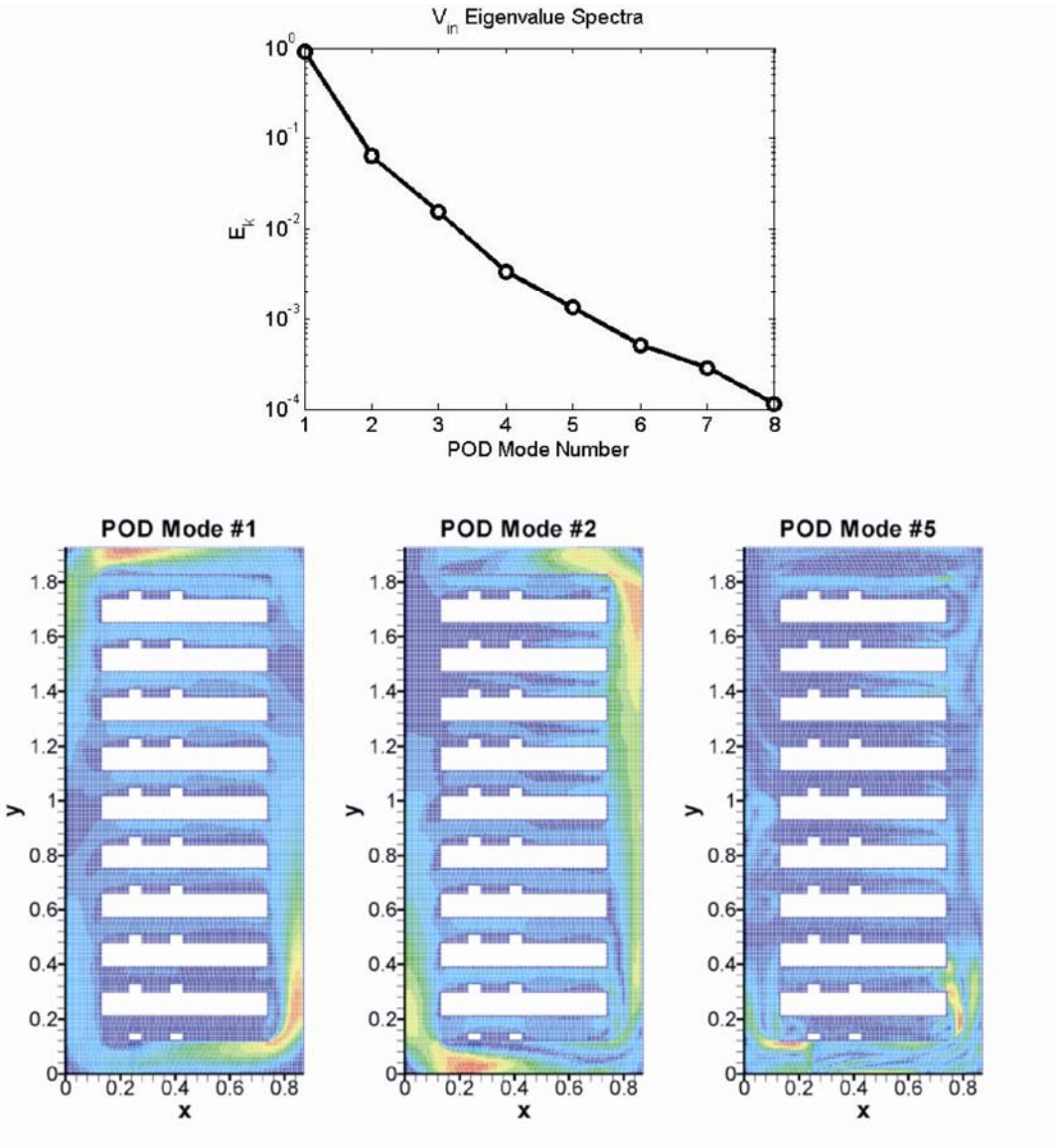


Figure 3.14. V_{in} eigenvalue spectrum and POD modes

The rapid decay of λ indicates that the first 2 POD modes capture the dominant modes of the system and the first 4 modes are able to reconstruct any *observation* with less than 0.001 L^2 error. Note that $\lambda \to 0$ with increasing mode number, indicating that the higher order POD modes do not contribute to the mechanics of the system and their computation may be laden with numerical error [56]. A numerical cutoff criterion for the minimum value of λ that produces meaningful modes is defined.

The results are demonstrated for 4 test cases. To approximate the solution for the test case not in the range of the observations ($V_{in}^{t} = 2.15$ m/s), the largest two values of V_{in}^{o} were used to construct $\vec{\phi}^{\perp}$. Figure 3.15 below compares the PODc eigenvalue spectrum with the original and shows the resulting u_{err} error measure.

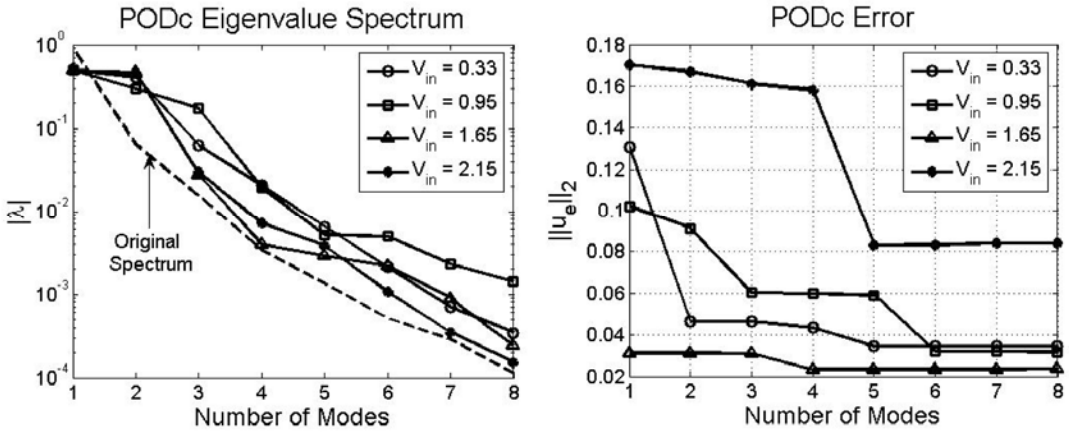


Figure 3.15. Single parameter spectrum and L^2 approximation error

The data show the error decreases monotonically as more modes are added and at about p = 5 modes, u_{err} has converged. For each test case, the desired inlet mass flux is matched accurately, $G_e = O(10^{-2}) \ \forall \ p$. The RMS error in velocity magnitude is less than 0.05 for all cases in the range of V_{in}^{o}, and only slightly larger for the case outside the range of V_{in}^{o}. The results also indicate that a majority of the error is incurred near the inlet and exhaust of the rack, while the flow inside the individual servers is accurately approximated. It is these regions of the flow field that have the most practical significance in the case of data processing cabinet airflow management.

3.7.2. Multi-Parameter RANS POD

Recall for the multi-parameter case, the inlet velocity is fixed at 0.5 m/s and 3 different server flow rates (0.08, 0.13 and 0.17 kg/s) are specified. The rack is assumed

to contain 3 low, 4 medium and 3 high flow rate servers to limit the design space. Various arrangements of the servers were solved with the full CFD/HT model for a total of m = 21 observations and additional configurations were solved to serve as test cases. Figure 3.16 shows the mean-centered POD eigenvalue spectrum and the first 3 POD mode shapes.

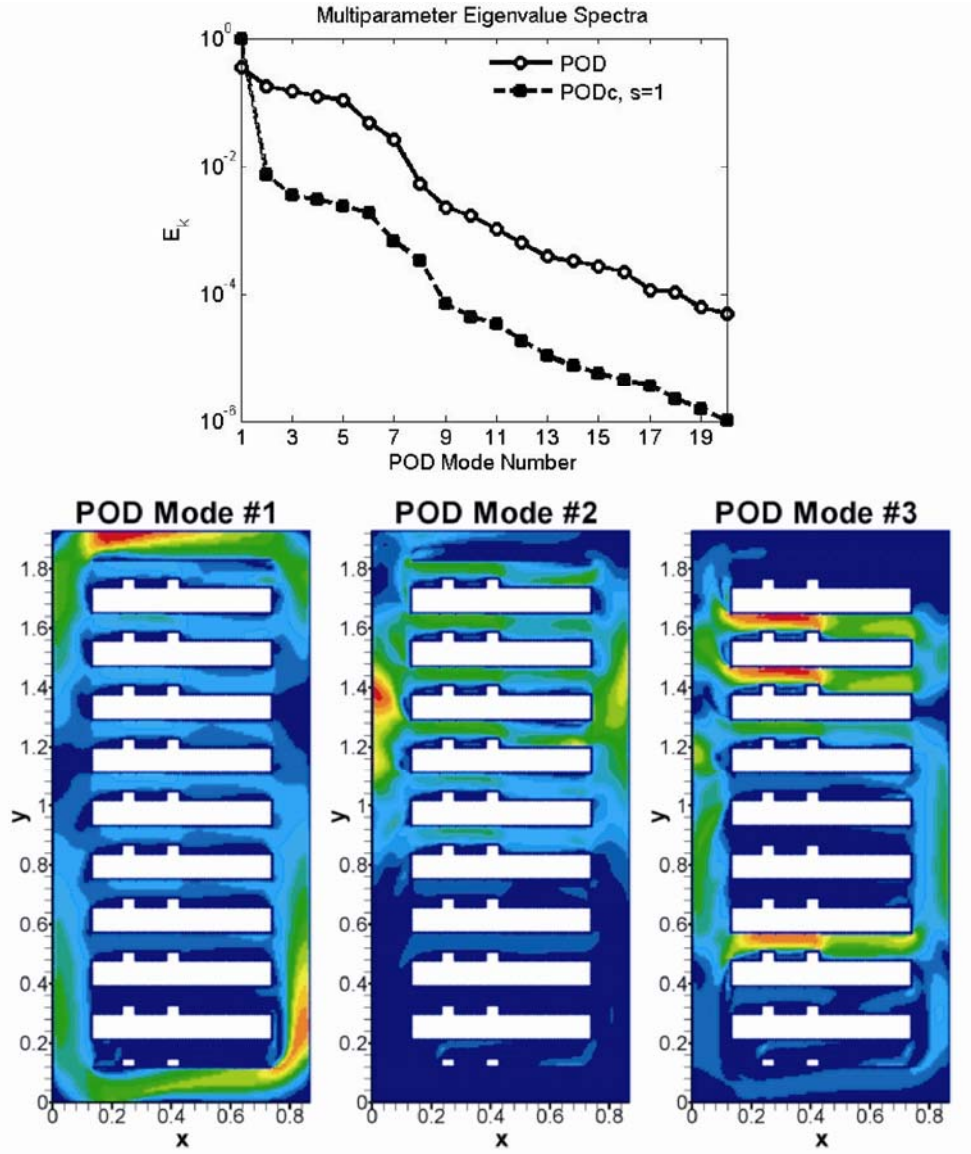


Figure 3.16. Multi-parameter POD modes

Each individual server's flow rate is unknown as it will deviate from the nominal value depending on its position due to the pressure field inside the rack. Thus, each observation is characterized by a vector containing 10 values of either '1', '2' or '3' for low, medium and high flow rate servers. Given an input vector, the mass of flow rates for a new approximate solution can be computed by taking a weighted average of the two nearest observations, as determined by (3.18).

Figure 3.17 plots the error measures for the ones of the test cases for various orthogonal complement subspace dimension, *s*. The data in Figure 3.17 show that using more observations to construct $\vec{\phi}^{\perp}$ will reduce the error for small *p*, but the errors become the equal as $p \to m-1$ independent of *s*. The flux matching procedure is able to satisfy the individual server mass flux goals with a relative error on the order of $O(10^{-2})$.

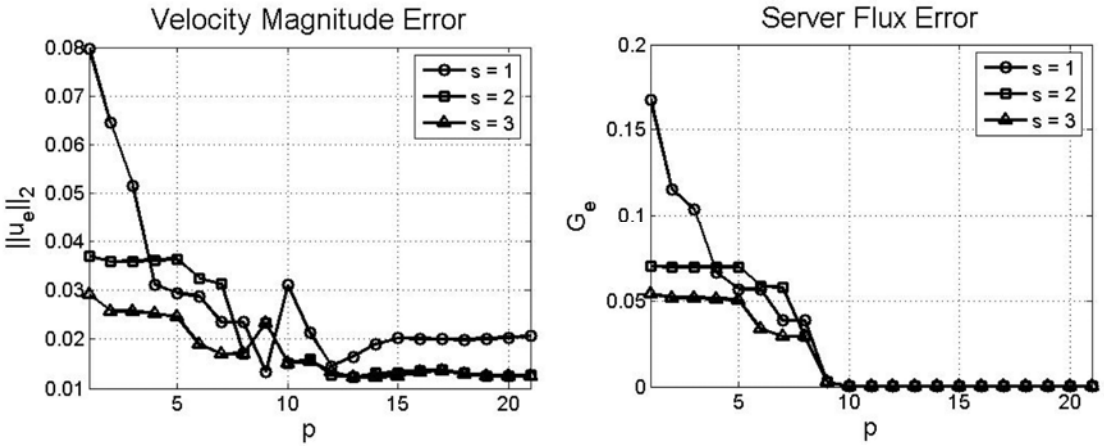


Figure 3.17. Multi-parameter PODc error

A final point to be made is that since the POD modes are themselves solutions to the governing equations, they predict the correct velocity profile into each individual server. Figure 3.18 plots the exact and approximate solutions bear the entrance to servers 7 and 8, where strong changes in mean strain rate result from recirculation effects. This is the primary reason why an integral term (mass flux) is used as the matching condition.

If a specific profile was to be matched, the desired profile might not correspond to an actual solution. By matching the flux, the correct mass flow rate is achieved and the detailed structure contained within the POD modes produce the local velocity field that corresponds to a physically realizable solution. This also implies that this method will produce the closest physical solution in the event that an unphysical goal is specified.

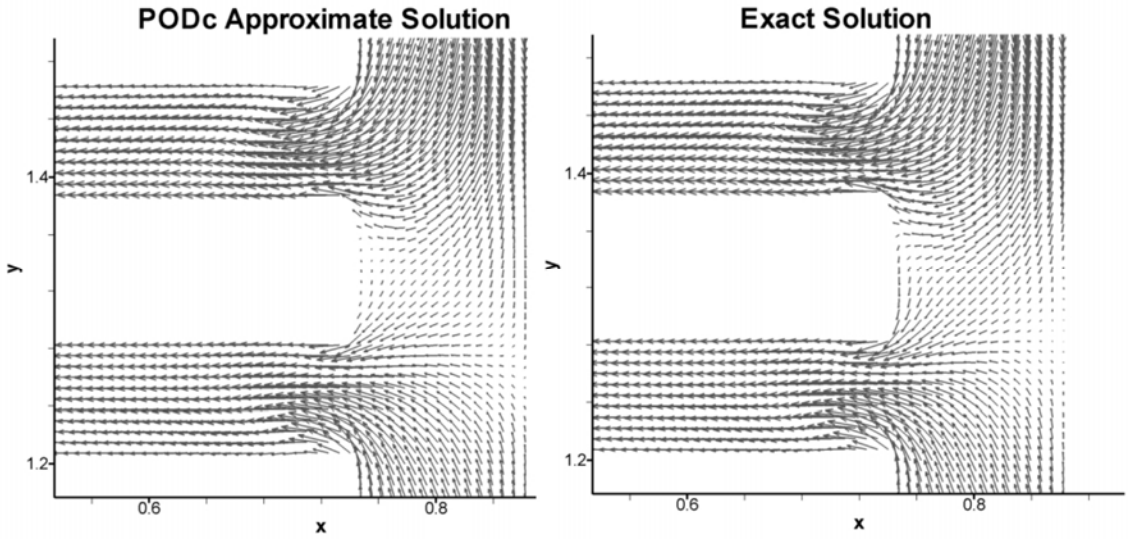


Figure 3.18. Local PODc and exact solutions near the entrance to servers 7 and 8

Figure 3.19 shows the exhaust region of servers 6 and 7 which is the region where the maximum error of all the test cases occurred. The PODc procedure is not able to completely resolve the sharp gradients that occur in secondary flows. But since the objective was to retain the dominant flow patterns for design type analyses, this tradeoff is more than acceptable.

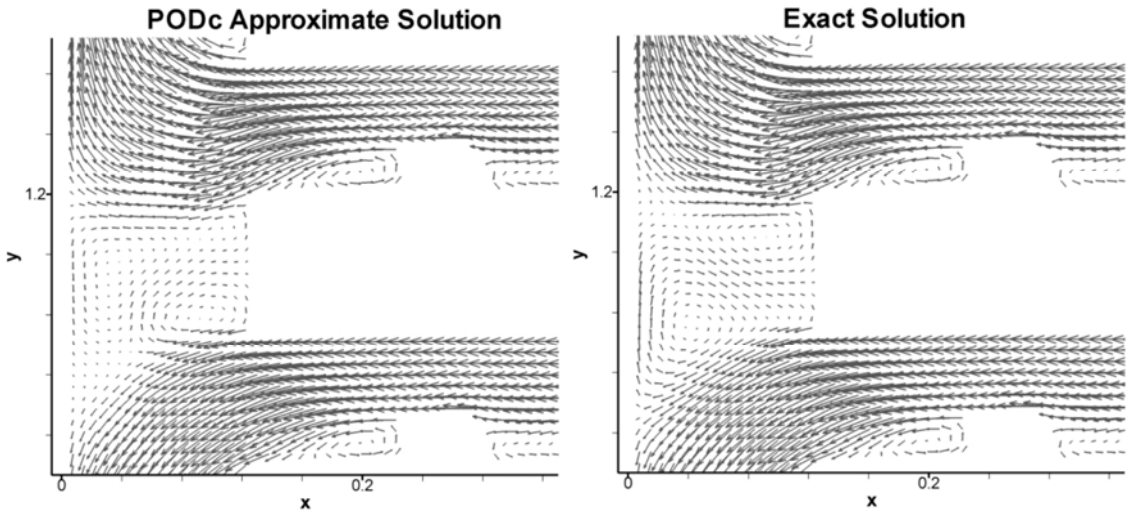


Figure 3.19. Local PODc approximations in region of maximum velocity error; exhaust of servers 6 and 7

3.7.3. Optimization

The reduced-order models produced by the PODc-FMP procedure allow designers of thermal systems to quickly assess candidate designs and perform optimization studies without cumbersome full-scale numerical models. To demonstrate this utility, robust design principles were used in conjunction with the reduced-order modeling methodology to determine the optimal inlet flow rate to the rack using the model and boundary conditions of §3.7.1. Robust design aims to not only minimize the value of the optimization function but also to minimize the curvature of the function such that small changes in the both controllable and uncontrollable design variables do not cause significant deviations away from the optimization point and result in system operation outside the feasible design space. Further details are provided in [65, 66] especially those concerning the application of robust optimization to design of electronics thermal management.

Using robust design principles, the air-cooled rack was optimized to dissipate 50% more power (2400 kW) and reduce the chip temperature variability 20% to 60%

depending on whether an optimal or robust configuration is employed [65, 66]. These robust design principles are accessible with reduced-order models where full-scale CFD/HT may be too time consuming to perform the analysis. The reduced-order flow model for the air-cooled rack and temperature solver have also be used in other optimization studies, including genetic algorithm based optimization routines [67].

The reduced-order model averaged 13 seconds per function execution. Figure 3.20 details the program flow and time required for each step. A majority of the time is spent assembling the linear system for the energy equation.

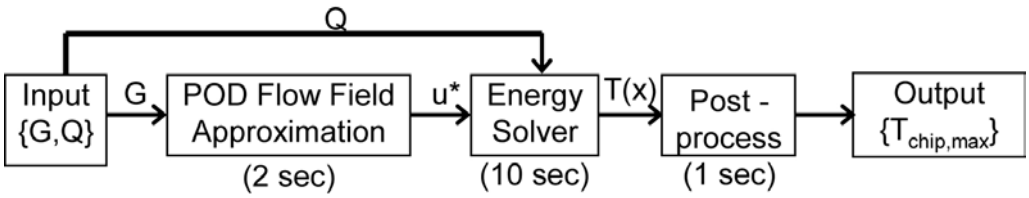


Figure 3.20. Reduced-order model execution and time requirements

Validation was performed by solving the full CFD model, which yielded chip temperatures within an average of 5 ^{o}C of the reduced-order computed solution. On a higher level of validation, the power distribution of the servers found to be most efficient yields an approximate hyperbolic tangent, demonstrated to be a highly efficient configuration in [29]. Even if very precise optimization points are required, the reduced-order model can be used with design optimization principles to compute a very good initial guess and then the full-scale model can be used to refine the optimization.

4. REDUCED-ORDER MODELING OF FORCED TURBULENT CONVECTION

The POD procedure developed in §3.4 provides a reasonably accurate and highly efficient method for constructing full field approximation of parametric turbulent flows and overcomes the difficulties associated with taking the Galerkin projection of the RANS equations without detailed knowledge of the effective viscosity. As seen in Figure 3.20, solving the temperature field given the reduced-order flow approximation creates a bottleneck in the rapid solution methodology. To overcome this, the flux matching procedure is extended to incorporate the RANS energy equation and special treatment is introduce to ensure the proper coupling of the velocity and temperature field in forced convection. The methodology is developed through an example of a prototypical forced convection situation in electronics thermal management.

4.1. Model Parameters

The methodology is illustrated for a RANS simulation of two-dimensional duct flow of air over two aluminum heated blocks in tandem (see Figure 4.1). The geometry is identical to the experimental measurements of Yoo *et al.* [68], whose data were used to validate the turbulence modeling in the CFD/HT code. The steady, incompressible, constant properties RANS continuity, momentum and energy equations without external forcing or buoyancy effects used to model the flow and heat transfer are:

$$\nabla \cdot \vec{u} = 0 \tag{4.1a}$$

$$\vec{u} \cdot \nabla \vec{u} - \nabla \cdot (\nu_{eff} \nabla \vec{u}) + \frac{1}{\rho} \nabla P = 0 \tag{4.1b}$$

$$\rho c_p \vec{u} \cdot \nabla T - \nabla \cdot (k_{eff} \nabla T) = 0 \tag{4.1c}$$

where $\nu_{eff} = \nu + C_\mu \frac{k^2}{\varepsilon}$ and $k_{eff} = k + \frac{c_p \nu_t}{\rho \Pr_t}$ with $Pr_t = 0.85$ and can be computed through any RANS-based turbulence model. The standard k-ε model with non-equilibrium wall functions [69] was used to model the effect of turbulence on the mean flow and the inlet velocity, turbulent kinetic energy and turbulent dissipation rate boundary profiles were calculated assuming a fully developed flow using Prandtl's $^1/_7$ power law. The fully converged CFD/HT model consisted of 10,000 grid cells and solving for two velocity components, pressure, k and ε at each grid cell resulted in approximately 50,000 DOF to model the flow. The solution was demonstrated to be independent of grid size and convergence criteria, and each solution required approximately 500 iterations to converge.

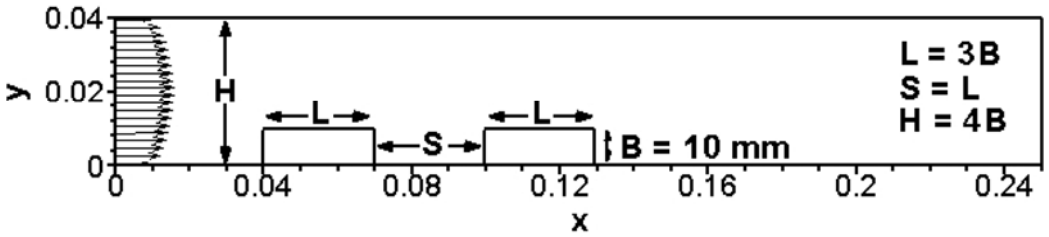


Figure 4.1. Model geometry from Yoo *et al.* [68]

The flow parameter range of the model was chosen to be $13{,}690 \leq Re \leq 41{,}070$, for $Re = H\bar{u}/\nu$ which corresponds to an average velocity of $5.0 \leq \bar{u} \leq 15.0$ m/s in air, and the block power was assumed to range from 25 to 200 W. The inlet temperature was fixed at 288 K and all fluid properties were evaluated at this temperature. All results will be reported as the temperature rise above this nominal value. Table 1 summarizes the set of observations used to construct the reduced-order model subspace.

Table 1. Turbulent convection observation database

k	Re	$\overline{u}$	Q_1 [W]	Q_2 [W]
1	13692	5	25	200
2	16430	6	50	175
3	19168	7	50	100
4	21907	8	100	150
5	24645	9	125	175
6	27383	10	100	100
7	30122	11	200	175
8	32860	12	100	75
9	35599	13	200	125
10	38337	14	200	50
11	41075	15	200	25

The heat input to each block was applied as a uniform heat flux on the bottom surface (y = 0). The local Nusselt number ($Nu = hB/k$), using a running coordinate over the surface of blocks, is plotted against the experimental data of Yoo *et al.* [68] in Figure 4.2. The numerical simulation agrees fairly well with the experimental data with some error in magnitude over the surface of the first block. Chen *et al.* [70] have experimentally investigated a similar geometry for similar Reynolds numbers and have suggested the standard low Reynolds number turbulence model of Jones and Launder [71] provides accurate local heat transfer coefficient predictions. The code is based on wall functions and the pressure-gradient sensitive wall functions employed provide the most accurate results without significant code modifications. It should also be noted that the mesh employed here is of similar size to that of Chen *et al.*, even though the wall functions are used in this investigation while the Jones-Launder low Reynolds number model relies on a damping function to model the near wall effects.

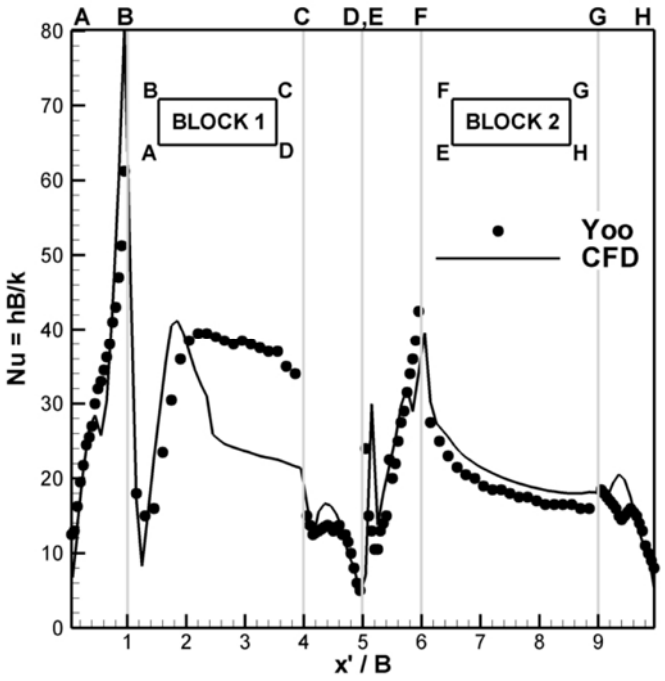


Figure 4.2. Comparison of numerical solution and experimental measurements at *Re* = 13,690, data from Yoo *et al.* [68]

The basic motivation of low-dimensional modeling is to create a more computationally efficient way of reproducing the physics described by a high fidelity numerical simulation or detailed experimental dataset. It is important to note that both numerical and experimental observations will depart somewhat from the true system behavior. The purpose of this investigation is to present a reduced-order modeling framework for turbulent forced convection, not to model a particular system. Thus, the numerical data will be treated as the 'exact' system response in the description of the methodology below, and some discussion of various errors and their contribution to total error will be subsequently provided.

4.2. Low-Dimensional Turbulent Flow Modeling

The orthogonal complement POD (PODc) with flux matching will first be demonstrated to approximate the velocity field for a randomly selected test case corresponding to Re = 36,320 ($\overline{u}^*$ = 13.33 m/s in air). The set of mass flux control surfaces reduces to a single surface coincident with the domain inlet (conversely, the domain outlet could be used to produce the same results by continuity). This simple flow has only a single parameter to be used as a matching condition, indicating the 2-term expansion $\vec{u} = \vec{u}_0 + a_1\vec{\varphi}_1$ is all that is available for the solution approximation. It has been demonstrated that the cumulative energy resolved by the first k modes produces an error bound on the approximation [72]. Figure 4.3 shows $E_1 \approx 0.98$, indicating the 2-term approximation should produce errors on the order of 2% in the sense of the L^2-norm. Equation (3.18) selects the observation k = 9 in Table 2, corresponding to $\overline{u}$ = 13.0 m/s, as the source function.

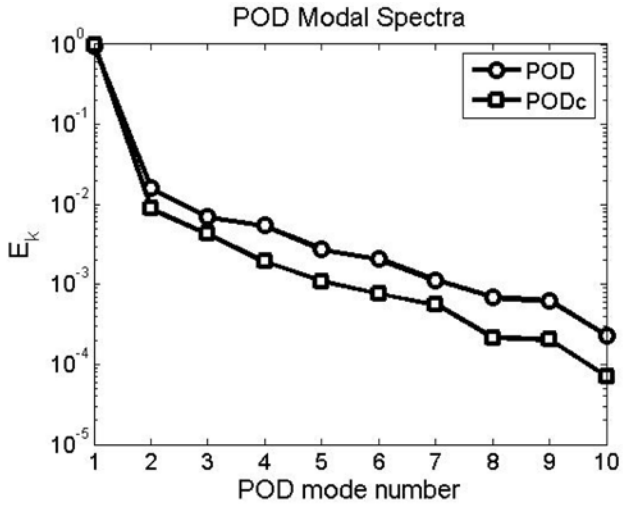


Figure 4.3. Mean-centered velocity POD and orthogonal complement POD (PODc) modal energy content

Figure 4.4a illustrates the decaying error in the velocity magnitude for the 1- and 2-term approximate solutions. Figure 4.4b plots the exact solution and the approximate velocity field in the vicinity of the leading edge of the first block, which is where the maximum error occurs. Both Figures 4.4a and 4.4b show that approximate solution is very accurate, especially considering the full CFD model requires 40,000 DOF to solve the flow and the reduced-order model contains only 2 DOF. The 2-term approximation captures the inlet velocity profile exactly, produces a maximum absolute point-wise error of 0.232 m/s and an L^2 – error norm over the domain of 0.003. Even though the observation space is dense with various mass flow rates, weighted averaging and scaling observations will generally result in poor approximations. For example, rescaling the observation with the difference in mass flow rate from the nearest observation ($\overline{u}$ = 13.0 m/s) produces errors 3 times as large as the PODc based approximation.

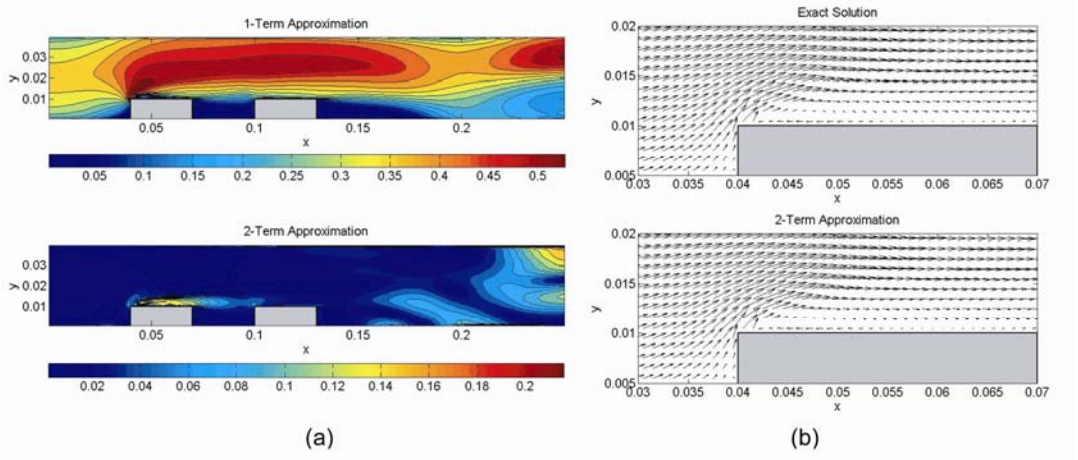


Figure 4.4. (a) Velocity absolute approximation error [m/s] and (b) detailed local velocity fields

At this point, it is reiterated that this reduced-order modeling procedure has been demonstrated on significantly more complex flows, comprised of multiple control surfaces, with very successful results [73] and the objective of the present study is to couple the energy equation into the methodology to extend the low-dimensional modeling framework to convective flows. As the system grows in complexity and more parameters are incorporated, more matching conditions are generated, which requires more modes to be retained in the approximation. Thus, the level of approximation keeps pace with growing system complexity.

4.3. Low-Dimensional Turbulent Convection Modeling

An efficient and accurate reduced-order modeling methodology for turbulent flows has been demonstrated in the previous section, and the objective is now to extend the procedure to include a low-dimensional solution to the energy equation. The orthogonal complement POD and flux matching procedure from the previous section will be employed because of their simplicity, and the main challenge will be in coupling the temperature and velocity fields. To begin, independent velocity and temperature decompositions will be assumed:

$$\vec{u} = \vec{u}_0 + \sum_i a_i \vec{\varphi}_i \quad and \quad T = T_0 + \sum_j b_j \psi_j \tag{4.2}$$

The temperature POD modes are computed with the same procedure, given the temperature observation matrix $T^{obs} = \{T_1, T_2, ..., T_m\} \in \mathbb{R}^{n x m}$.

A natural way to couple the velocity and temperature fields is the Galerkin projection, but as with the v_{eff} term in RANS momentum equation, the RANS energy equation would require k_{eff} to be specified. Thus, substituting the approximate velocity

field $\vec{u}^*$ into (4.1c) and projecting onto the subspace spanned by $\Psi = \{\psi_1, \psi_2, ..., \psi_m\}$ is considered ineffective here.

The flux matching procedure (FMP) will be extended to include the energy equation, accordingly the heat flux function can be defined analogous to (3.11) as:

$$F_h(T) = \int_{\Gamma_h} -k \frac{dT}{d\hat{n}} dx \tag{4.3}$$

The heat flux control surfaces correspond to the 3 surfaces of each block exposed to the airflow. Alternatively, the bottom surface of each block where the heat flux is applied could also be used as the control surface because the system is steady.

POD modes are solutions to the governing equations (4.1a-c) and inhomogeneous modes can be viewed as a solution with arbitrary boundary conditions. The flux functions (3.11 and 4.3) define an inverse problem of finding the corresponding boundary conditions. When the flux function involves a gradient, approximation with discrete data can produce large errors, especially if the gradient is sharp relative to the measurement point spacing. This can be especially difficult if the observations were generated through CFD/HT data, where wall functions were used to alleviate near-wall grid resolution requirements when integrating the turbulence transport equations. Temperature wall functions based on $T^+(y^+) \equiv (T_w - T_p)\rho c_p u_t / q_w^{"}$ are used to link the wall boundary condition to the temperature in the first grid cell, T_p. When generating the observations either T_w or $q_w^{"}$ is specified, but in the temperature POD modes, only T_p is known, rendering the evaluation of wall heat flux or temperature an under-determined problem.

The wall fluxes can be evaluated by recalling that the method of snapshots expresses the POD modes as a linear combination of the observations. This can generally be written as:

$$\vec{\Phi} = L\vec{U} \tag{4.4}$$

where the linear transform L involves a SVD operation. If the POD procedure is thought of as finding the principle axes of the data contained in $\vec{U}$, then L can be thought of as the m x m rotation matrix between $\vec{U}$ and $\vec{\Phi}$. The transformation matrix L can be computed as the projection of the observations onto the modes utilizing the pseudo-inverse again and the modal fluxes can be directly computed, *viz*:

$$L = \vec{U}^{+}\vec{\Phi} \tag{4.5a}$$

$$F_m = (v^T L)^T \tag{4.5b}$$

The vector v defines the observation mass fluxes, $F_m(U^{obs})$, and the transpose operation in (4.5b) is to maintain the same dimension between F_m and v. Defining the temperature observation matrix and the associated matrix of block heat inputs, $Q = F_h(T^{obs})$, the modal heat flux can be computed as:

$$F_h = (Q^T T^{+} \Psi)^T \tag{4.6}$$

This procedure can be used to evaluate any flux function that defines the same quantity contained in the goal vector G regardless of where Γ is located in the domain.

A common method of treating coupled phenomena is to work with an extended state vector, 'ESV', see [48]. Defining $Y = \begin{bmatrix} \vec{u}^* & T \end{bmatrix}^T$, the modal expansion is modified to:

$$\begin{bmatrix} \vec{u}^* \\ T^* \end{bmatrix} = \begin{bmatrix} \vec{u} \\ T \end{bmatrix}_0 + \sum_{i=1}^{p} a_i \begin{bmatrix} \vec{\varphi} \\ \psi \end{bmatrix}_i \tag{4.7}$$

and the goals of boundary conditions are concatenated to $G = \begin{bmatrix} G_m & G_h \end{bmatrix}^T$. The mass and heat flux functions are also concatenated to form the extended modal flux function $F = [F_m(\vec{\varphi}) \quad F_h(\psi)]$, leading to the ESV optimization problem to find the weight coefficients,

$$\min\{ \left\| F(\vec{u}^*, T^*) - \sum_{i=1}^{p} a_i F(\vec{\varphi}_i, \psi_i) \right\| \} \ , \ k = 1,2,...m \tag{4.8}$$

which can be solved by the goal residual technique of (3.13a-c). Since the weight coefficients must simultaneously account for the velocity and temperature goals, a slower rate of converge is expected, implying more modes will generally be required to satisfy the boundary conditions. This could be a major shortcoming in parametric system model reduction where observations are expensive to generate. Note that the reduced-order flow model uses the ESV method to couple the velocity components, $\vec{U} = \begin{bmatrix} \vec{u} & \vec{v} \end{bmatrix}^T$, otherwise the divergence free condition would be violated.

Solutions based on the ESV method will not be demonstrated because poor results should be expected as satisfying both flow and thermal conditions simultaneously imposes conflicting restrictions on the weight coefficients. This can be seen by considering the separate decompositions (4.2) and operating the mass and heat flux functions on the velocity and temperature modes. The vector of modal mass fluxes has an order of magnitude of $|| F_m || \sim 10^{-3}$ and the modal heat fluxes are $\left\| F_h \right\| \sim 10$, using either the POD or PODc procedure. The mass flux goal is $\left\| G_m \right\| \sim 10^{-1}$ kg/s and the heat

transfer rate goal is $\|G_h\| \sim 10^2$ W. Noting the order of magnitude of a matrix and its pseudo-inverse are related by $\|F^+\| \sim \|F\|^{-1}$, the modal weight coefficients are respectively $\|a\| = \|F_m\|\|G_m\| \sim 10^2$ and $\|b\| \sim \|F_h\|\|G_h\| \sim 10$. Thus, trying to combine mass flow and temperature goals will result in poor approximations and rescaling either velocity or temperature modal fluxes (i.e working in units of kW instead of W) will greatly increase the error in other.

A new method for coupling the temperature field to the velocity field needs to be devised. Ideally, the temperature source function would be a solution to (4.1c) with $\vec{u}^*$ as the velocity field and corresponding k_{eff}. Then, the linearity of (4.1c) for a known k_{eff} could be used to rescale the solution as $T_0 = cT(\vec{u}^*)$. An approximation to this would be to 'borrow' the temperature field associated with $\vec{u}_0$ and use the POD modal expansion to perturb the solution until the boundary conditions are satisfied. The source function can also be scaled to improve the approximation since it is treated as the dominant mode of a linear system:

$$T_0 = cT(\vec{u}_o) + \sum_{i=1}^{p} b_i \psi_i \tag{4.9}$$

Reasonably accurate temperature field solutions may possibly be obtained without implicitly coupling the velocity field, however this would disregard (4.1c), possibly producing unphysical results and lack the rigor needed for a robust methodology intended for more complex flows. The implicit coupling and source function scaling introduce no additional complexities implementing (4.9) algorithmically. The same PODc procedure of (3.15-17) is employed on the temperature observations with:

$$\Psi = \psi_0 + \psi' \quad , \quad where \;\; \psi_0 = T(\vec{u}_0(G_m)) \;\; and \;\; \psi' \in R^{n\,x\,m-1} \tag{4.10}$$

To scale the source function properly, concatenate $\psi = [\psi_0 \quad \psi']$ and apply the sequential flux matching procedure of (3.13a-c) as:

$$\Delta G_{h,i} = G_{h,i-1} - F_h(T_{i-1}) \tag{4.11a}$$

$$b_i = F_h^+(\psi_i)\Delta G_{h,i} \tag{4.11b}$$

$$T_i = \sum_{j=1}^{i} b_j \psi_j \tag{4.11c}$$

4.4. Results

To demonstrate the reduced-order temperature solution, the previous test case of Re = 36,320 ($\vec{u}^* = 13.33$ m/s) will be used with an arbitrary power dissipation $G_h = [Q_1 \quad Q_2] = [96 \quad 66]$ W. Figure 4.5 plots the POD and PODc temperature modal spectra and it can be seen that the PODc produces a slightly steeper spectrum in the lower order modes. This is a favorable property, as resolving more of the dominant physics with fewer modes allows one to truncate the expansion (4.9) earlier for a given accuracy requirement. It is generally the case that the temperature spectrum decays less sharply than the velocity spectrum, as noted by other researchers [31], implying more temperature modes are required for the same order of accuracy as the velocity approximation. Figure 4.6a illustrates a few basic mean-centered POD modes and Figure 4.6b shows the PODc modes, both for a section of the domain near the heated blocks surface where the largest temperature gradients occur. The PODc procedure produces different temperature modes than the POD procedure because the PODc formulates the modes as a function of parameter values contained in *G*, converting $\psi(x) \rightarrow \psi(x, G_h)$.

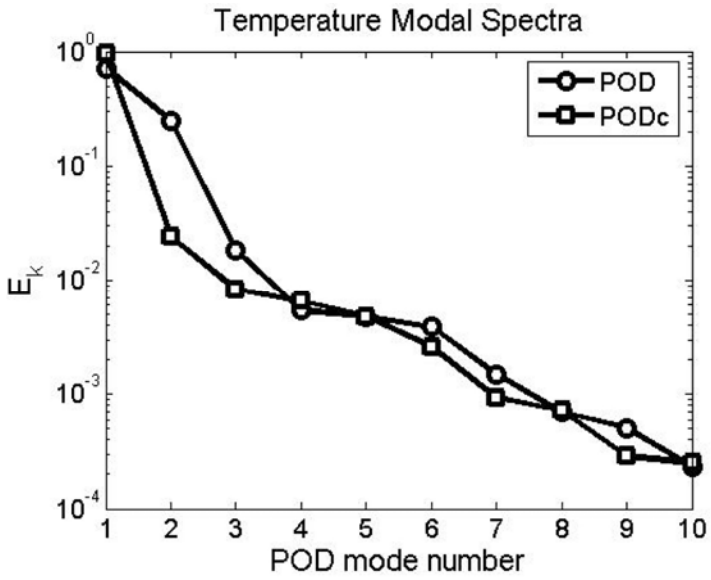


Figure 4.5. Temperature modal spectra for the POD and PODc procedures

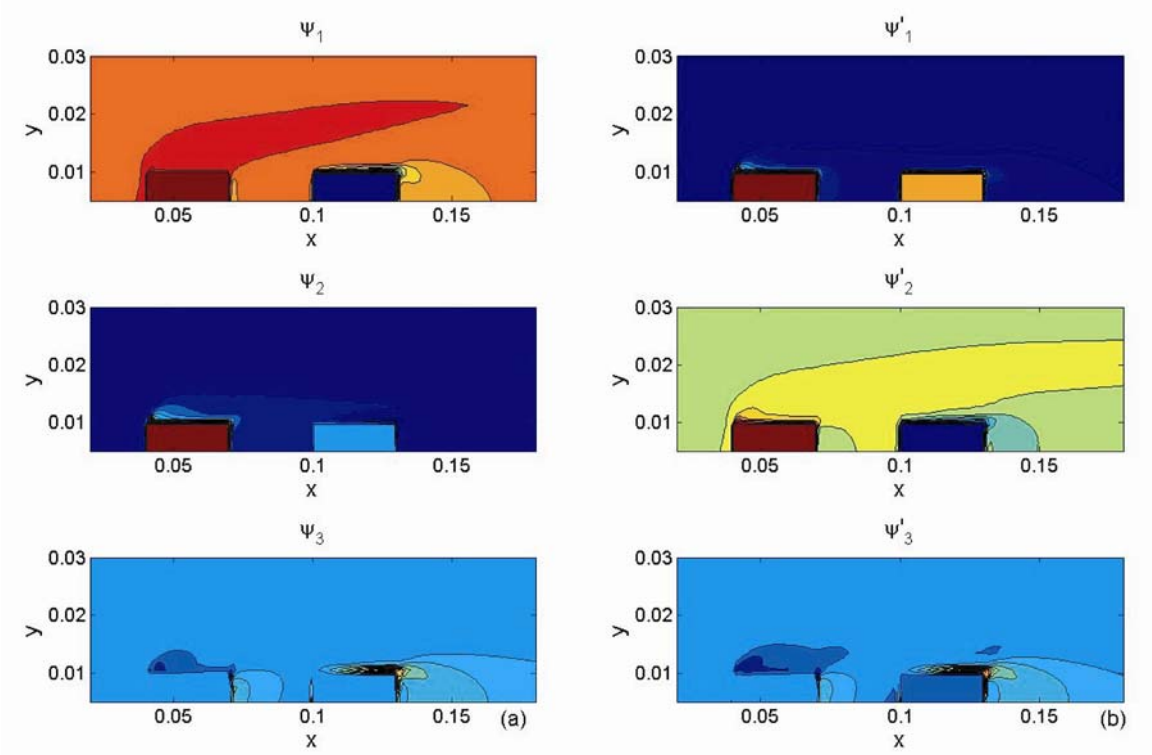


Figure 4.6. Local (a) mean-centered POD modes and (b) PODc modes for test case

Using all p = 10 modes, the implicit coupling has a maximum point-wise error of 0.529 °C. The relative error over the domain and boundary condition satisfaction is defined as

$$T_{err,k} = \frac{\left\| T_k^* - T_{exact} \right\|_2}{\left\| T_{exact} \right\|_2} \tag{4.12a}$$

$$Q_{err,k} = \frac{\left\| F_h(T_k^*) - F_h(T_{exact}) \right\|_2}{\left\| F_h(T_{exact}) \right\|_2} = \frac{\left\| F_h(T_k^* - T_{exact}) \right\|_2}{\left\| F_h(T_{exact}) \right\|_2} \tag{4.12b}$$

and the implicit coupling procedure with the PODc subspace and flux matching method of evaluating the weight coefficients produced T_{err} = 0.0322 and Q_{err} = 4.17 x 10^{-4} . Figure 4.7 illustrates the exact and approximate solutions. The largest errors occur near the surface of the blocks where the largest temperature gradients occur. A large truncation in system DOF can allow the dominant physics to be captured, but at the expense of some small-scale features being discarded, usually in the form of sharp gradients. However, these errors are of the order of 3%, making the temperature approximation quite accurate considering the system was reduced to 10 DOF from the original 60,000 DOF required to compute the turbulent flow and heat transfer. The approximate solution shows a slightly overly diffusive temperature field near the trailing edge of both blocks, which may be partially attributed to a finite error between the desired and approximate solution boundary conditions.

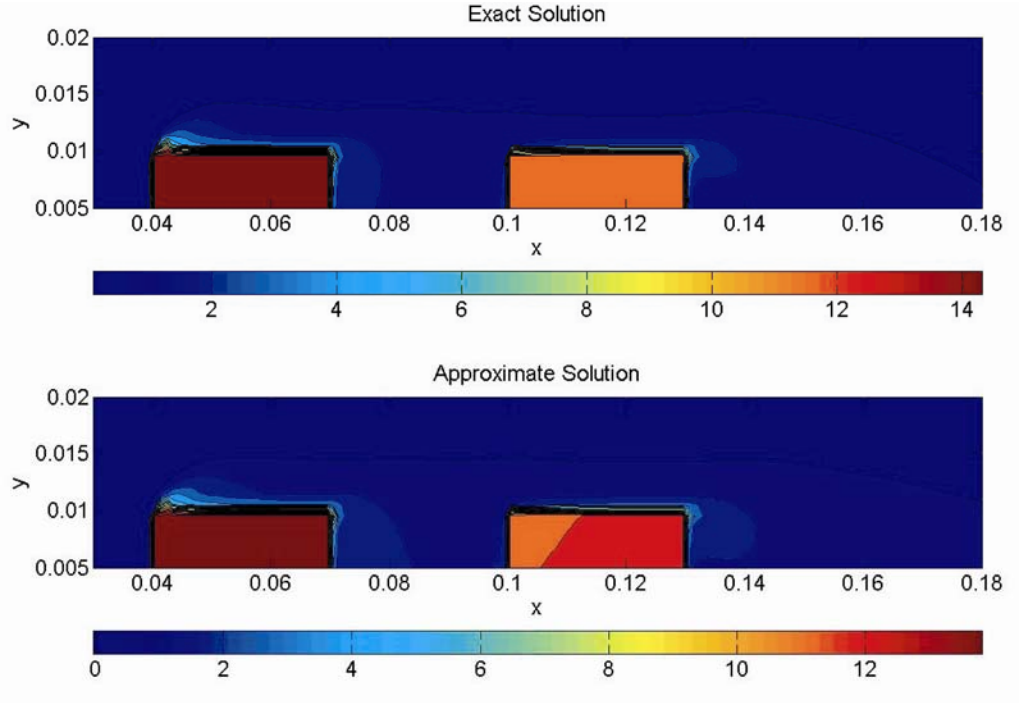


Figure 4.7. Exact and approximate temperature fields [°C]

Both velocity and temperature low-dimensional models are constructed by using a linear subspace to describe the physics for a range of parameters. Poor approximations will result for the nonlinear RANS momentum equation (4.1b) if the POD or PODc procedures are used outside the parameter range, however the linearity of (4.1c) with known $\vec{u}^*$ and k_{eff} allows one to predict a temperature field from any parameter value as long as $\vec{\Phi}$ and Ψ subspaces adequately describe the physics. If the boundary heat fluxes were large enough to induce significant buoyancy or even phase change in the case of a liquid medium, the Ψ subspace would not describe the thermal physics, for instance. For an inlet Reynolds number of Re = 23,221 ($\bar{u}$ = 8.48 m/s) and block power dissipation of $G_h = [Q_1 \quad Q_2] = [349 \quad 475]$ W, the maximum point-wise temperature error was 3.07 °C (out of a maximum of 127 °C) and T_{err} = 0.0227. The integral boundary condition formulation was satisfied to $Q_{err} \sim 10^{-5}$. The approximate velocity field had a 0.48 m/s

maximum error and a relative L^2 error of 0.0043. Note the observation data in Table 4.1 ranges from $1/8 \leq Q1/Q2 \leq 8$, so it is reasonable to expect that any forced convection flow within the parameter range of Re and Q_1/Q_2 would perform with comparable accuracy.

5. ERROR ANALYSIS

In the POD methodology, the question of the minimum number of modes to be retained in the reduced-order model often arises. The Galerkin projection produces *m*-coupled ODE's in time for the weight coefficient evolution and reducing the number of equations to be integrated in time can result in significant economies for long term dynamics investigations. The objective of this work was to produce accurate steady models using the minimum number of system observations. In either case, it must be demonstrated that the POD subspaces ($\vec{\Phi}, \Psi, \Pi, \ldots$,) sufficiently capture the system physics. Some authors use projection energy of the un-retained POD modes as a total error estimate $\|e\| = 1 - E_{1\to m} = \sum_{k=p+1}^{m} \lambda_k \,/\, \sum_{k=1}^{m} \lambda_k$, although this assumes there is no in-plane error [72]. Note $\|\cdot\|$ denotes the 2-norm throughout this section unless otherwise noted. Christensen *et al.* [56] have suggested using the distance in the POD subspace between the approximation and the ensemble mean as an error measure. The PODc was developed to specifically correct this limitation in the classical POD methodology and new error estimates need to formulated. The parametric modeling methodology is based on a low number of system observations, leading to a relatively few number of basis functions and generally requiring that most, if not all, modes will be retained. It must be noted that the number and distribution of observations cannot be determined *a priori* for a given nonlinear system and is currently the largest concern and limitation in the POD methodology [38].

5.1. Error Partitioning

Rathnam and Petzold [63] divide the error into the subspace projection error (e_o) and in-plane error associated with evaluating the modal weight coefficients (e_i). Figure 5.1 sketches the POD subspace as the optimal linear fit to an ensemble of data and schematically depicts this error partitioning for a general POD subspace Φ. The term u^{obs} is the system observation to be approximated, u^* is the approximate POD solution and u^p is the affine orthogonal projection of u^{obs} onto Φ and represents the best POD approximation of u^{obs}. To show Φ contains sufficient information, the *a posteriori* error estimate of $\|e_o\| \approx 0$, or at least $\|e_o\| << \|e_i\|$, can be used to show the error is dominated by the in-plane contribution.

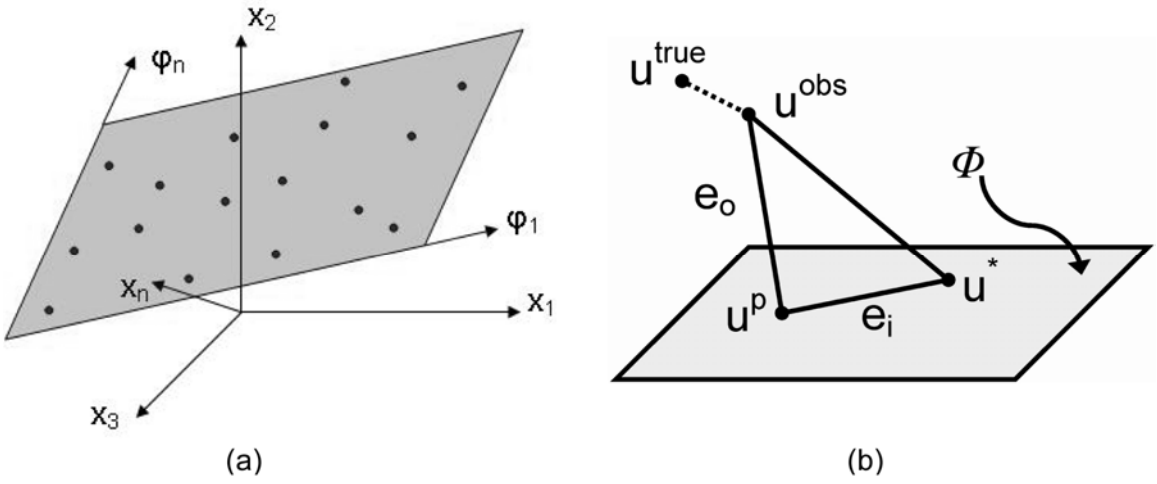


Figure 5.1. a) POD subspace and b) in-plane (e_i) and out-of-plane (e_i) error components

The error between the observation and the 'true' solution will not be considered as it is the user's task to ensure that numerical or experimental data faithfully represent the true system. The POD approximate solution (u^*) is an efficient solution to the full model (u^{obs}) and only describes the physics contained within Φ. Thus, the low-

dimensional model will generally not be more accurate that the full model in the sense of being closer to the true solution, but it can produce nearly as accurate results as the full model in an exceedingly more efficient manner.

To examine the convergence of the sequential solution procedure of (3.13a-c) and (4.19a-c), a dual weighted residual technique [74] will be used. Consider the canonical non-square optimization problem:

$$\min\left\{\left\| G - \sum_i a_i F_i \right\|\right\} \rightarrow Fa = G \tag{5.1}$$

This could be solved directly in a least squares manner as $a = F^+ G$ where $F^+ = (F^T F)^{-1} F^T$ is the matrix pseudo-inverse. The POD modes are normalized and ordered in descending projection energy so the modal weight coefficient magnitude should generally decay. Computing the a_i*'s* sequentially will mimic this spectral decay because the goal residual will decrease with each successive mode while computing the vector of a_i*'s* all at once as $F^+ G$ does not guarantee this decrease in coefficient magnitude.

Define *d* as the vector of distances between the approximate weight coefficients (a^*) and the projected weight coefficients ($\tilde{a}$) in the modal subspace. The true projected and approximate solutions to (4.13) are then:

$$F\tilde{a} = G \quad (\textit{true projected solution}) \tag{5.2a}$$

$$Fa^* = G \quad (\textit{approximate solution}) \tag{5.2b}$$

The error functional is defined as $J(a) = d^T a$, resulting in an error of $J(\tilde{a}) - J(a^*) = J(e) = (e, d)$ and a residual of $r = G - Fa^*$. The boundary condition error in state space (*e*) is analogous to the in-plane error (e_i) in the POD subspace. Note a

small residual does not imply a small error. The dual problem can then be formulated as a linear problem driven by the error functional, see (5.3), and the error functional can be expressed:

$$F^T a' = d \tag{5.3}$$

$$J(e) = (e, d) = (e, Fa') = (F^T e, a') = (r, a') \tag{5.4}$$

The fourth term in (5.4) was derived from the third term using Lagrange's identity $(u, Kv) = (\hat{K}u, v)$ where $\hat{K}$ is the adjoint of *K*, which reduces to K^T for $K \in \mathbb{R}$. From (5.4) it can be deduced that the error for the i^{th} term in the sequential solution is bounded by [74]:

$$\left|J(e)\right|_i \leq \sum_{k=1}^{i} \left(\left|r_k\right| \left\|a'_k\right|\right) \tag{5.5}$$

This estimate provides an *a posteriori* error bound because knowledge of the vector of weight coefficient perturbations from the true projected solution is required.

5.2. Error Estimates

The POD subspace representation of the physics has been demonstrated to be adequate and the sequential flux matching procedure has shown to be bounded and weakly converges to the desired boundary conditions. However, the overall objective is to demonstrate how well the reduced-order model approximates the full model. Consider a full nonlinear steady model in the canonical form $N(u) = 0$, $u \in \mathbb{R}^n$. The reduced-order canonical solution of the form $u = u_0 + \sum_{i=1}^{m} a_i \varphi_i$ can be expressed in the arbitrary POD subspace $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_m\}$ through the action of the orthogonal projector *P* from §3.3

as $\tilde{u} = P^T(u - u_0) + u_0$, $P \in \mathbb{R}^{n \times m}$. The full ($u$) and reduced-order approximate (u^*) solutions can be written as [63]:

$$u = P^T x + (P')^T z + u_0 = P^T x - e_o + u_0 \tag{5.6a}$$

$$u^* = P^T x + P^T y + u_0 = P^T x + e_i + u_0 \tag{5.6b}$$

where P' is the orthogonal complement to P. As above, the approximations will employ all m POD modes such that $\|e_o\| \approx 0$ rendering $u \approx \tilde{u}$. Solving for the in-plane error $e_i = P^T y$ and substituting in the modal expansion gives:

$$\begin{aligned} y &= \left(P^T\right)^+ (u^* - P^T x) = \left(P^T\right)^+ (u^* - \tilde{u}) \\ &= \left(P^T\right)^+ \left(\sum_i a_i^* \phi_i - \sum_i \tilde{a}_i \phi_i\right) = \left(P^T\right)^+ \sum_i \phi_i (a^* - \tilde{a})_i \end{aligned} \tag{5.7}$$

It can be seen that the projection matrix is simply the concatenation of the POD modes, $P = \Phi = \{\varphi_1, \varphi_2, ..., \varphi_m\}$, such that $\Phi^T U^{obs} = a$. Since Φ is an orthogonal matrix, $\Phi \sum_{i=1}^{m} \varphi_i = I \in \mathbb{R}^m$, and (5.7) reduces to:

$$y_i = a_i^* - \tilde{a}_i = -d_i \tag{5.8}$$

Comparison of the reduced-order approximation to an affine projection of the full model on the POD subspace produces the same error estimate as the adjoint problem to the boundary condition error. Therefore, the boundary condition error bound can be viewed as an in-plane error estimate $\|e_i\|$, suggesting the implicit assumption of matching only the relevant control surface integral conditions in the flux matching procedure (FMP) is a valid approximation. Again, this assumes that the out plane error is negligibly small, $\|e_o\| \to 0$ relative to the in-plane error. As a final note, the triangle inequality can be used

to bound the total error $\|e_o + e_i\| \le \|e_o\| + \|e_i\|$ even though $\|e_o\| << \|e_i\|$ for parameter values within the predefined range.

Error bounds are generally difficult to establish for POD-based models because the basis functions are problem dependent. The introduction of a flux function to evaluate the weight coefficients in lieu of standard Galerkin methods poses additional hardship in establishing error bounds because of the definition of the flux function of (3.11) depends on the parameters of the specific model. However; a heuristic *a priori* error estimate is available because of the rapidly decaying contribution from each successive POD mode. The principle of the orthogonal complement subspace decomposition is to allow the source function to satisfy a dominant portion of the parametric inhomogeneity and use the remaining POD modes to perturb the approximation to satisfy the residual on the control surface. Assuming that the contribution of the POD modes to the approximation is less than the source function, the following *a priori* error estimate is available:

$$\|u_0 - u_{exact}\|_2 > \|u_k^* - u_{exact}\|_2 \quad where\ u_0 = u^{\perp} \tag{5.9}$$

The assertion is that $u^{\perp}$ dominates the approximation in the sense $\|u^{\perp}\|_2 > \|u_k^* - u^{\perp}\|_2 \ \forall k$, which is generally satisfied by the application of (3.18) to construct the orthogonal complement subspaces. The POD methodology has a tendency to localize errors, which can be attributed to the POD subspace not completely resolving all the physics of the full model. The 2-norm is used to quantify the error over the entire domain and the subscript will be dropped throughout the remainder of this section for brevity. The kth approximate solution is given by $u_k^* = u^{\perp} + \sum_{i=1}^{k} a_k \varphi_k$, where the flux matching procedure determines

the k[th] weight coefficient as $a_k = F^+(\varphi_k)\Delta G_k$, see (3.13). The weight coefficient is thus on the order of $|a_k| \sim \left\|F^+(\varphi_k)\right\|\left\|\Delta G_k\right\| = \left\|F(\varphi_k)\right\|^{-1}\left\|\Delta G_k\right\|$ and the contribution of each successive mode is then $\|u_k\| \sim |a_k|\|\varphi_k\| = \|F(\varphi_k)\|^{-1}\|\Delta G_k\|$. The term $\|\Delta G_k\|$ exhibits decay by the application of the flux matching procedure and the magnitude of $\|F(\varphi_k)\|^{-1}$ is fixed because the POD modes are normalized. This causes a decay in $|a_k|$ so the contribution of each successive to the total approximation, and hence the incurred error, decays. These arguments are heuristic in nature because of the definition of the flux function, but holds for all applications of the methodology demonstrated here.

5.3. Turbulent Flow Error

The rack model introduced in §3.7.2 will be used to demonstrate the error associated with the PODc and FMP reduced-order modeling methodology using the *a posteriori* error analysis of the previous section. The first step is to validate the POD subspace by determining the out of plane error by considering the relative error between norm of the observation (test case) and the norm of the test case projected onto the POD subspace. This is computed as $\|e_o\| = \left(\|\vec{u}^{\,p}\| - \|\vec{u}^{\,obs}\|\right)/\|\vec{u}^{\,obs}\| = 0.0644 = O(10^{-2})$ showing the subspace constructed by the PODc procedure adequately represents the full model flow characteristics. This is consistent with the boundary condition error from §3.7.2, which was also shown to be $O(10^{-2})$.

Figure 5.2 compares the weight coefficients obtained by projecting the test case onto the POD subspace define the true projected solution, denoted with the subscript 'true' with the weight coefficients obtained by the FMP. The modal weight coefficients

obtained by the FMP are very close for the first 10 modes and the variations for the remaining modes do contribute a significant amount of error because their magnitude is small relative to the first few dominant modes of the system. The FMP values for the weight coefficients for $p > 10$ all have the correct sign and the magnitude error is due to the mode shape not producing a significant mass flux through some of the control surfaces. The modal flux is relatively small for these modes and the weight coefficient is artificially enlarged as the FMP attempts to satisfy the mass flux residual. The classes of flows considered for this methodology involve closed domains with multiple inlets and outlets and the resulting POD mode shapes do not generally exhibit this recirculation behavior near the control surfaces. Small recirculation regions may exist in the POD modes, but this behavior is relegated to much higher order modes where the mass flux residual is already very small. The error in weight coefficient evaluation may also be partially attributed to a nonzero out of plane error component.

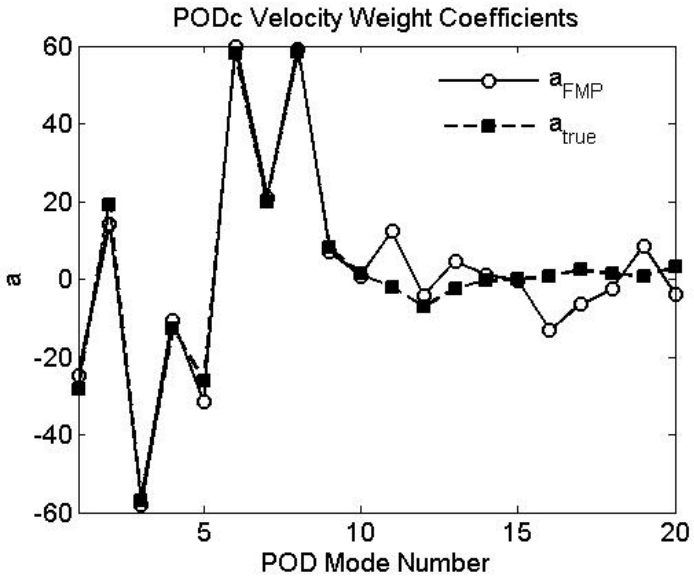


Figure 5.2. Velocity weight coefficients computed by the FMP and true values obtained by projection onto the POD subspace

The error bound estimate for the test case was $O(10^2)$ for each successive mode. This may not appear to be a particularly sharp error bound, but it does imply the FMP coefficients are of the correct order of magnitude, implying the overall velocity approximation is bounded.

In relation to the *a priori* error bounds of §5.2, the flux function is on the order of $\|F(\varphi_k)\| \sim 10^{-3}$ and goal residual decays from ~10^{-1} to ~10^{-2}. The weight coefficients are then ~10^2, rendering at most $\|u_k\| \sim 10^2$. The source function is $\|u^{\perp}\| \sim 10^3$ and thus represents an error bound on the ROM approximate solution.

5.4. Turbulent Convection Error

For the approximate temperature solution to the test case presented in section §4.4, the out of plane error was $\|e_o\| = \|T^p\| - \|T^{obs}\| = O(10^{-4})$, consistent with the heat flux boundary error, implying the temperature POD modes adequately capture the system physics. Figure 5.3 plots the true projected and FMP-computed weight coefficients for the temperature field approximation. The temperature modal weight coefficients demonstrate the same behavior of closely matching the first few dominant modes and then show small deviations for the higher order modes. The resulting approximations benefit form the strong decay in weight coefficient magnitude from $O(10^3)$ to $O(10^{-3})$.

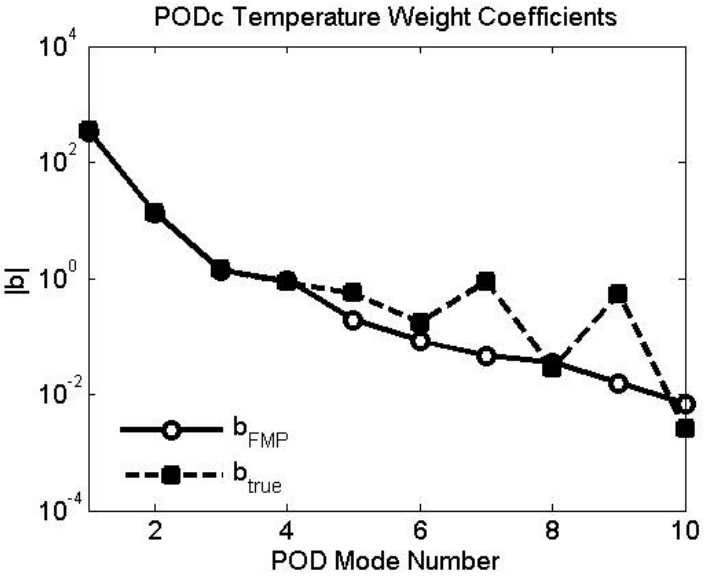


Figure 5.3. Temperature weight coefficients computed by the FMP and true values obtained by projection onto the POD subspace

Figure 5.4 plots the boundary condition matching error and the error bound estimate of (5.8) for the temperature field. The error bound is sharp and decays with increasing modes, indicating that the PODc-FMP approximate temperature solution is weakly convergent for forced turbulent convection.

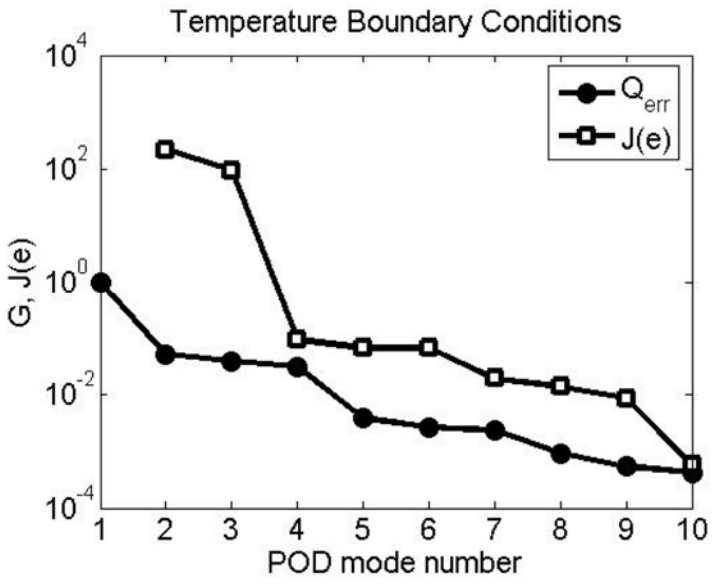


Figure 5.4. Temperature flux matching procedure error and error bounds

6. INTERCONNECTED REDUCED-ORDER MODELS FOR MULTISCALE DOMAINS

Fluid and thermal transport processes occur across a range of length scales in systems-level electronics thermal management. The primary scales of interest span the individual server to the largest length scale of cooling scheme when modeling racks and data centers. Such domains are considered multiscale as they contain a series physically separated domains, such as the individual servers inside a rack and the individual racks inside a data center. Characterization and modeling of theses systems are challenging because each rack inside a data center facility may contain a variety of different servers and each server may contain different electronic components.

An efficient strategy to bridge length scales is to develop separate models for various system components and assemble them together to model the full system. From an analysis and design viewpoint, modularity is a key benefit as various subsystem models can be integrated to form a full system in order to investigate their interactions. Decomposing a system into subsystems can greatly improve model efficiency by using various levels of description for different components. Complex engineering systems often involve transport processes across several length scales and a single computational model would require a large number of grid cells to stretch across these length scales. Modularity affords the ability to quickly integrate new components into an existing model without developing a new computational grid for altering the subcomponent models. It should also be noted that system decomposition can be applied to any system and there is not inherent error associated with treating a system as an interconnected set of subcomponents, [32].

6.1. Individual Component Modeling

For a representative example of systems-level thermal management of electronics, consider a series of channels connecting 2 plena, where each channel represents a server and the plena are representative of front and rear plena of a data processing cabinet. These systems are modular in nature, consisting of a series nested sub-domains (see Figure 6.1). The ability to maintain efficient models of the individual components and connect them together to evaluate various designs would result in a very efficient analysis tool.

To demonstrate the methodology, consider a highly simplified 2-dimesional model of an air-cooled rack containing a few servers. The model will be constructed from reduced-order model (ROMs) of an intake plenum and the repeated use of a single server model, as is shown in Figure 6.1. The plenum measures $2L$ x $5L$ and the server measures $4L$ x L, with L = 0.1 m, and both models were developed for the range $6{,}900 \leq Re_L \leq 31{,}300$. Table 2 lists the observations used to construct the component ROMs.

In lieu of specifying the boundary velocity, the observations may be constructed with pressure boundary conditions and the flux function of (3.11) can be used to compute the boundary mass flux and locate the observation in parameter space of G. Specifying pressure boundary conditions may be less intuitive and more difficult to generate observations with a specified Re range, but is an equally applicable method of generating observations.

To the best of the author's knowledge, the only works developing POD-based flow models with pressure effects are that of [53] which redefined the induced norm to model compressible flow, [59] which introduced a linear correction for the pressure drop

through a channel and flows driven by distributed pressure-velocity momentum sources to model fans was introduced by [73].

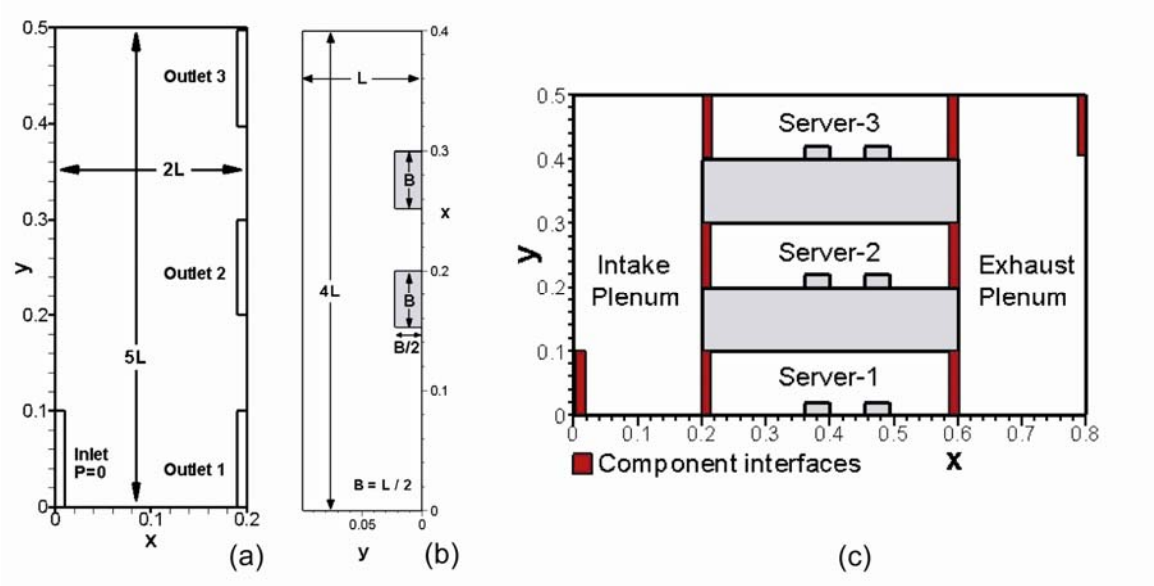


Figure 6.1. a) Plenum b) server component model geometry and c) full-scale system

Table 2. Plenum flow observations, ΔP [Pa] and $\dot{m}$ [kg/s]

	Outlet 1		Outlet 2		Outlet 3	
k	ΔP_1	$\dot{m}_1$	ΔP_2	$\dot{m}_2$	ΔP_3	$\dot{m}_3$
1	-0.865	0.135	-7.739	0.066	-6.862	0.145
2	-21.930	0.637	-189.040	0.492	-180.560	0.532
3	-5.930	0.287	-73.493	0.258	-83.699	0.499
4	-15.729	0.599	-63.445	0.281	-57.049	0.147
5	-3.709	0.228	-54.420	0.455	-42.748	0.137
6	-17.084	0.606	-93.345	0.173	-100.250	0.477
7	-11.067	0.489	-63.583	0.179	-62.927	0.333
8	-3.602	0.237	-50.177	0.296	-42.878	0.257
9	-16.335	0.606	-79.506	0.229	-71.647	0.275
10	-17.217	0.608	-104.950	0.484	-89.191	0.174
11	-9.315	0.363	-117.290	0.418	-117.020	0.495
12	-8.939	0.361	-113.730	0.444	-106.570	0.430

Server Flow Observations

k	1	2	3	4	5	6
ΔP	1.500	7.200	12.900	18.600	24.300	30.000
$\dot{m}_1$	0.136	0.297	0.400	0.481	0.550	0.611

Results for the intake plenum sub-model will briefly be presented to illustrate the dominant flow features. The velocity and pressure modal spectra for the plenum and server models for representative test cases are plotted in Figure 6.2. The use of the pressure modal spectrum will be explained shortly.

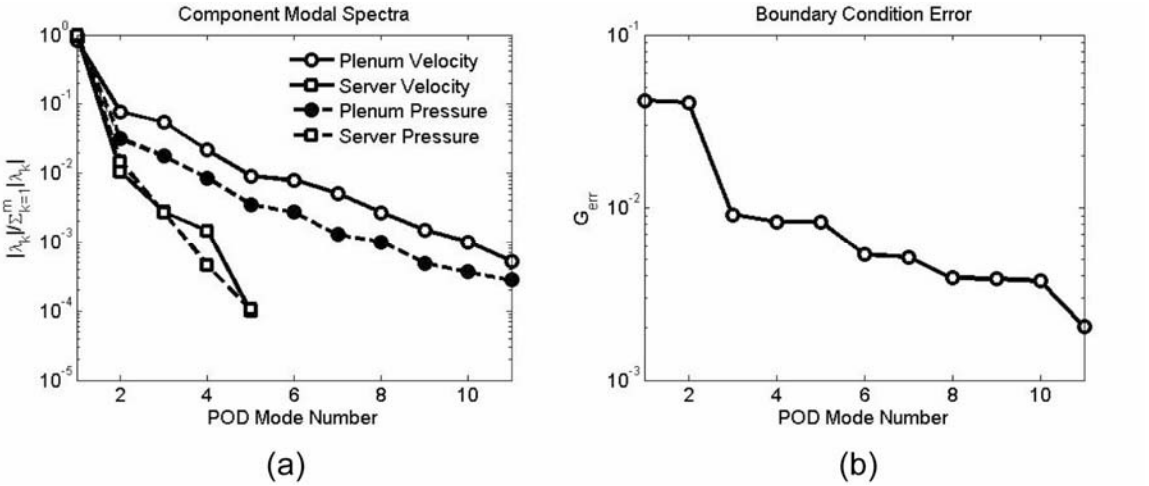


Figure 6.2. a) Server and plenum sub-component modal spectra and b) intake plenum boundary condition error

The relative L^2 error, defined again as $u_{err,k} = \left\|\vec{u}_k^* - u_{exact}\right\|_2 / \left\|u_{exact}\right\|_2$ for the velocity field and likewise for the boundary conditions, showed that the boundary conditions were satisfied to with *3%* and velocity approximation error was 6.2%. Error contours for a test case corresponding to a mass flux of G = {0.2070, 0.1473, 0.3602} kg/s are plotted in Figure 6.3. The plenum model contained 4000 grid cells, or 20000 total DOF to model the flow considering *u, v, P, k, ε* are solve for in each grid cell, while the reduced-order model contains only 12 DOF, for an $O(10^2)$ reduction in DOF.

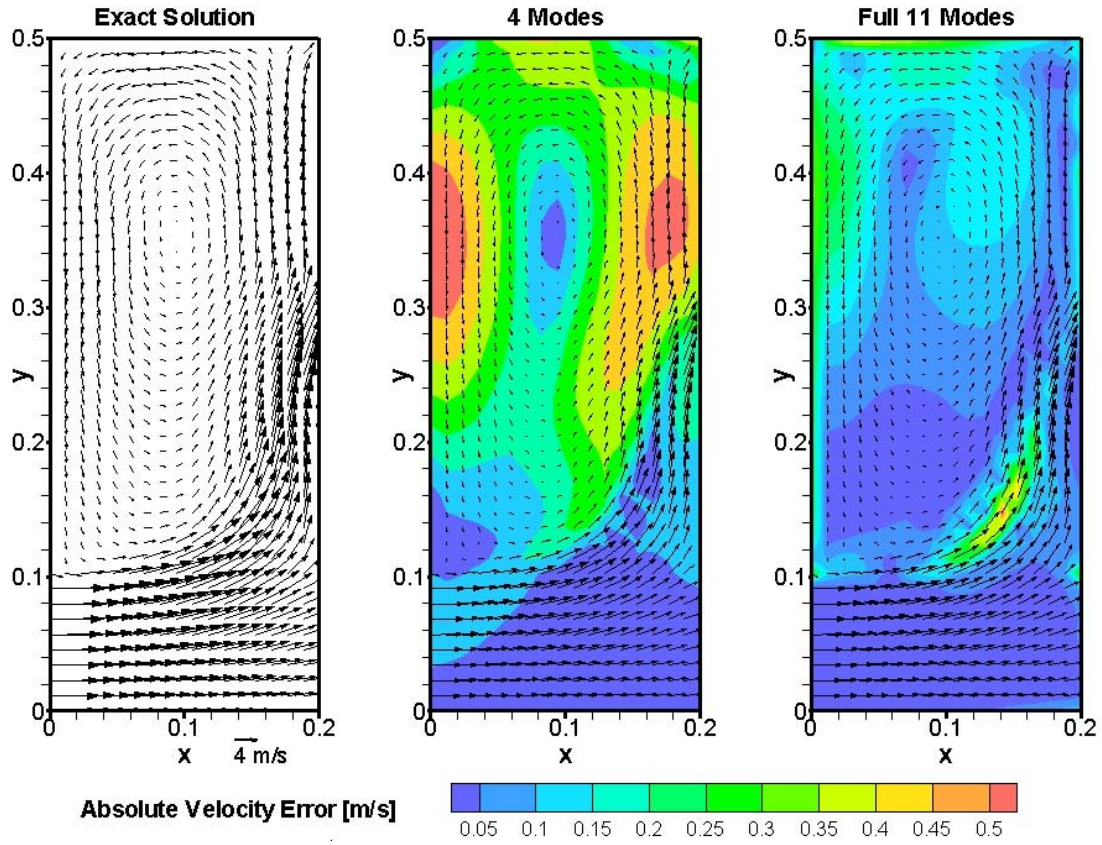


Figure 6.3. Intake plenum error field

6.2. Pressure Field Approximation

The ROMs do not need to be autonomous because the input mass flux can be computed with a full-scale CFD simulation or be derived from another ROM. To couple the ROMs, it must be noted that specifying the component interface mass fluxes *a priori* decouples the ROMs because there is no driving force between the two components and they behave essentially independent from one another. Many flows in electronics cooling are pressure-driven, such as fans moving air through a series of vents and channels, and it is the pressure difference across these components that drive the flow. A much-celebrated property of POD analysis is the elimination of pressure for incompressible

flows; however, coupling of ROMs for the class of flows presented here requires knowledge of the pressure field.

The ROM pressure fields can be approximated by using the PODc procedure to construct a pressure modal subspace, $\Pi = \{\pi_1, \pi_2, ..., \pi_m\}$. The eigenvalue spectra of Π strongly follow the velocity modal spectra, as is expected due to the role of pressure in incompressible flows. Using the velocity modal weight coefficients from the PODc-FMP produces accurate isobars shapes, however the magnitude of the field is poorly scaled. The velocity coefficients based pressure approximation is reasonably accurate up to a constant, i.e. $P^* = c\left(P(\vec{u}^{\perp}) + \sum_i b_i \pi_i \right)$, the unknown constant cannot be determined.

To approximate the pressure field accurately, the pressure POD modes are projected back onto the observations ensemble $P^{\text{obs}} = \{P_1, P_2, ..., P_m\}$ to obtain the set of observation weight coefficients,

$$b^* = \Pi^+ P^{obs} \in R^{m x m} \tag{6.1}$$

A G-dimensional quadratic response surface of the form $b_k^* = f(G^{obs})$ is computed for the k^{th} pressure mode as a function of the observational mass fluxes, G^{obs}. The k^{th} weight coefficient is then evaluated as $b_k = f(G^{obs})$ and the approximate pressure field is assembled as

$$P^* = \sum_i b_i \pi_i = \sum_i f_i(G^{obs}) \pi_i \tag{6.2}$$

The response surface methodology can be applied one mode at a time because the pressure modes are linearly independent and b_k are uncorrelated. Regression methods are the only available tool to compute b because the ROM are parametrized by the boundary

mass flux and no pressure information about the new approximate solution is available to use in a matching procedure. Figure 6.4 below plots the velocity weight coefficient and response surface approximations for a test case using the intake plenum component ROM.

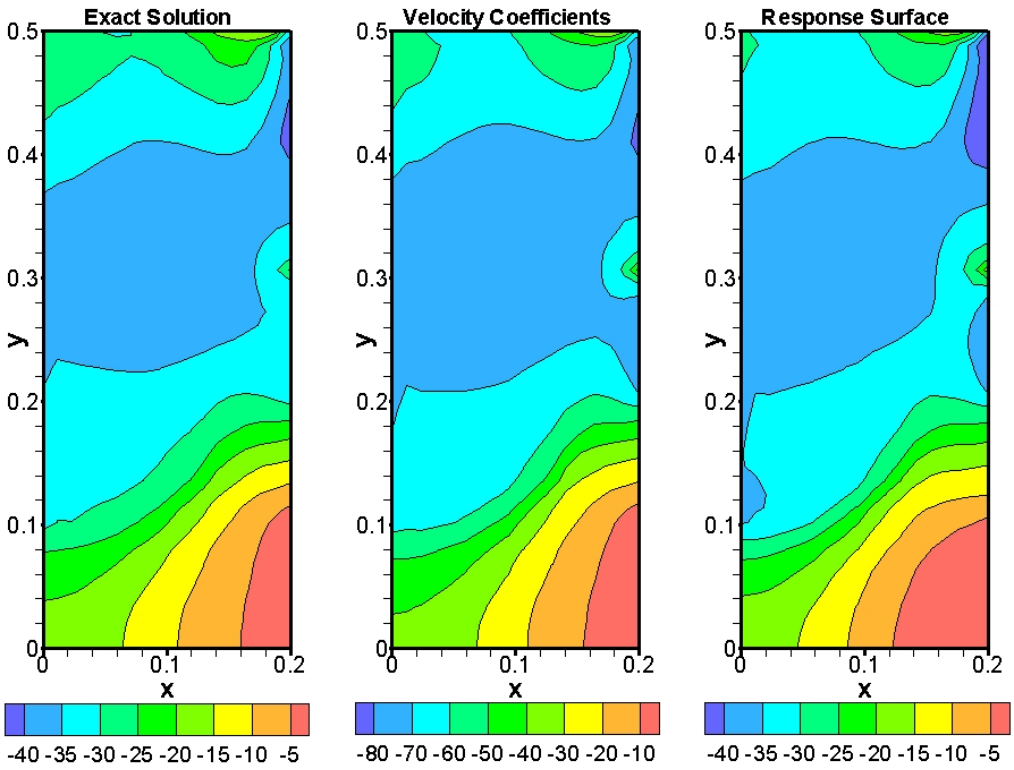


Figure 6.4. Component ROM pressure field approximation

With the response surface methodology producing satisfactory results for the pressure field, it would be natural to consider the same solution process to determine the velocity weight coefficients. The immediate issue is that the boundary conditions are not enforced in the weight coefficient computation and one relies on the assumption of the response surface accurately approximating the parametric conditions as well as the solution over the entire domain. Figure 6.5a compares the first couple velocity weight coefficients using the FMP and response surface (denoted 'rs') approximations and Figure 6b plots the boundary condition error using the response surface method.

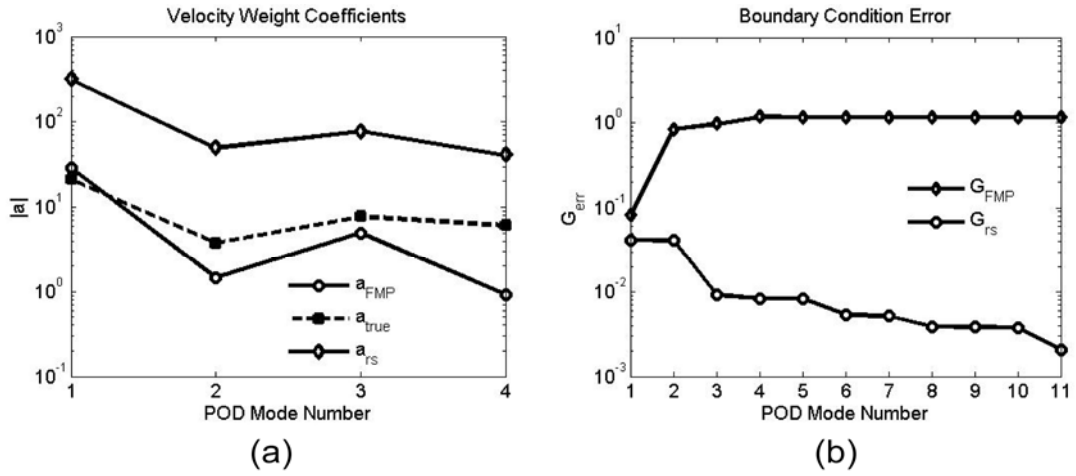


Figure 6.5. Response surface approximation for a) weight coefficients and b) boundary condition approximation error

Figure 6.6 plots the pressure weight coefficients and compares the values against the true weight coefficients obtained by projecting the pressure test observation onto the POD pressure modal subspace, $b_{true} = \Pi^{+} P^{obs}$. The reason the response surface method produces acceptable results for the pressure coefficients and not the velocity coefficients is that pressure coefficients tend to be larger in magnitude and exhibit a strong spectral decay so that error in the response surface predictions are less amplified in the final field approximation relative to the velocity approximation. Note this strong spectral decay was also exhibited by the scalar temperature field in §5.4

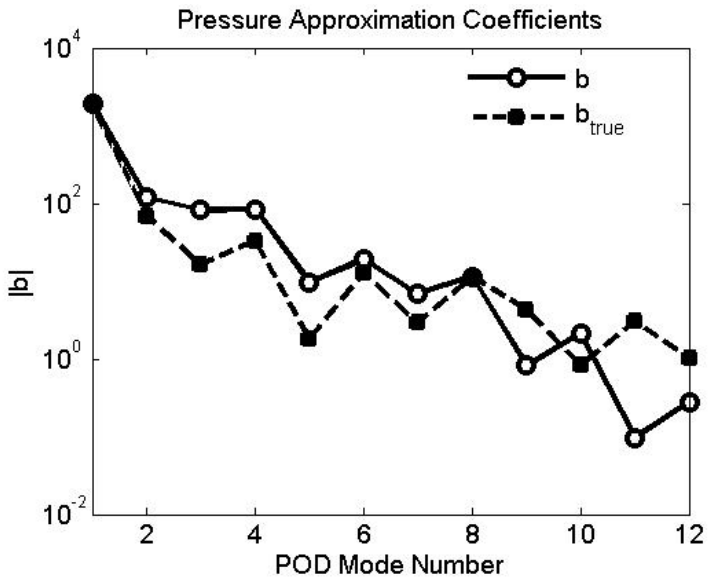


Figure 6.6. Comparison of response surface and true pressure weight coefficients

6.3. Systems-Level Modeling

The system-level model is constructed by joining the 3 server ROMs to the inlet plenum and the exhaust plenum, which is a mirror image of the intake plenum both in the *x*- and *y*-directions, will collect the outlet air from the servers and exhaust it to the ambient. To drive the flow, fan models will be placed at the outlet of the server models to mimic an induced fan and lumped resistance models are included between the intake plenum and sever inlet to model the pressure drop across an inlet vent. The fan is modeled with the cubic pressure-velocity model (S_f) and the lumped resistance model (S_s) is representative of inertia losses in high *Re* flows. The inlet and outlet pressures to the full system can be assumed to be zero without loss of generality. The source functions for the lumped parameter models are evaluated using area-averaged velocity $\overline{u}$:

$$p(\overline{u}) = 200 - 40\overline{u} + 20\overline{u}^2 - 4\overline{u}^3 \equiv S_f(\overline{u}) \tag{6.3a}$$

$$p(\bar{u}) = 15.3125\bar{u}^2 \equiv S_s(\bar{u}) \tag{6.3b}$$

A new POD basis needs to be computed for the exhaust plenum, even though the exhaust plenum has the same geometry because the flow observations are highly nonlinear and reversing the flow through the plenum produces significantly different flow patterns. The exhaust plenum ROM was developed from the same observations as the intake plenum with the mass fluxes specified in the opposite direction from the inlet to the outlet. It will be noted that the ROM geometry is fixed because geometric scaling of POD modes is a source of ongoing research and beyond the scope of this investigation.

Each component sub-domain will be identified with a superscript as Ω^j, j = 1,..,5 , and the kth control surface (interface) for the jth sub-domain in (3.11) is termed Γ_k^j, providing the G_k^j mass flux. The minimization of (3.12) is altered to:

$$\min\left\{\left\| G^j - \sum_i a_i^j F^j(\vec{\varphi}_i^j) \right\|\right\} \quad \text{where } F^j(\vec{\varphi}_i^j) = \int_{\Gamma_i^j} \rho \varphi_i^j \cdot \hat{n} dx \tag{6.4}$$

Continuity can be used to reduce the number of unknown ROM fluxes to G_{k-1}^j and the ROMs do not need to be solved concurrently to ensure satisfaction of (6.4). Concepts from flow network modeling (FNM) [75] are used to generate the matching conditions between the component models and assemble the full system. FNM calculates the pressure at specific nodal locations and the flow between two nodes is driven by momentum equations of the form $P_1 - P_2 = S(\bar{u})$. This technique produces acceptably accurate results when the momentum equation models conduit-type components, such as pipes, channels or even the servers in this example because the flow resistances can be expressed as $\Delta P_{1-2} = A\bar{u} + B\bar{u}^2$. For manifold-type flows, such as the intake plenum of Figure 6.1, $\Delta P_{i-1} = f(\Delta P_{i-2}, \Delta P_{i-3})$ and no simple relations are readily available from

flow resistance handbooks. Accurate pressure field approximations are available from the ROMs and pressure drop between nodes can be computed by integrating the pressure over the control surfaces. The system nomenclature and FNM pressure locations are illustrated in Figure 6.7.

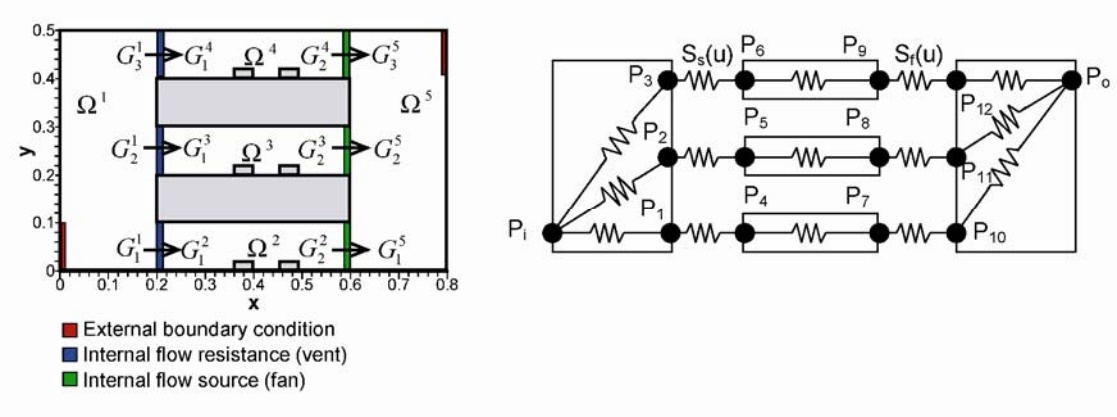


Fig 6.7. Full-scale system nomenclature and flow resistance network

The standard SIMPLE algorithm [76] is used to solve for the nodal pressure and momentum link flow rates with the only alteration being that the momentum source is evaluated using the ROM pressure field. The procedure is completely analogous to pressure-velocity coupling methods in incompressible CFD and will be outlined here with more details available in [75]:

FNM – SIMPLE Algorithm

1. Guess a nodal pressure distribution P
2. Use the momentum link equations $\Delta P = S(\bar{u})$ to calculate the momentum link flow rates given the nodal pressures

3. Use the nodal continuity equation $\sum_i \rho \bar{u}_i A_i = \sum_i G_i = 0$ to combine the momentum link equations into a linear system and solve for a new pressure distribution
4. Return to Step 2 and repeat until convergence

In general, the momentum equation will be nonlinear in u and an appropriate linearization must be introduced to perform step 3 above. This source term linearization, also equivalent to standard CFD methods, can be obtained by expressing the momentum source term as constant plus the linear variation in u:

$$\Delta P = C + Lu \tag{6.5}$$

The constant and linear variations are obtained by taking a Taylor series about the previous iteration:

$$S(u) \approx S_{i-1} + \left.\frac{\partial S}{\partial u}\right|_{i-1} (u_i - u_{i-1}) \tag{6.6a}$$

$$C = S_{i-1} + \left.\frac{\partial S}{\partial u}\right|_{i-1} u_{i-1} \quad and \quad L = \left.\frac{\partial S}{\partial u}\right|_{i-1} \tag{6.6b}$$

6.4. Full-Scale Results

Using the SIMPLE procedure to assemble the above system showed that the relative L^2 error norm for the approximate solution mass flux over all interfaces was 0.0316 and the relative L^2 error norm over pressure nodes P_1 to P_{12} was 0.063. Figure 6.8 plots the true and approximate velocity and pressure fields. The pressure fields in the various sub-domains can be properly scaled after the FNM solution calculates the nodal pressure distribution in Figure 6.7. For example, P_4 is known and $P_7 = P_4 - S(\Omega^2)$

where $S(\Omega^2)$ is obtained by integrating the pressure over the inlet and outlet of Ω^2. The linearity of pressure can than be used to shift the pressure field $P(\Omega^2) = P^*(\Omega^2) + P_1 + P_4$ relative to the full-scale model reference pressure $P_i = 0$.

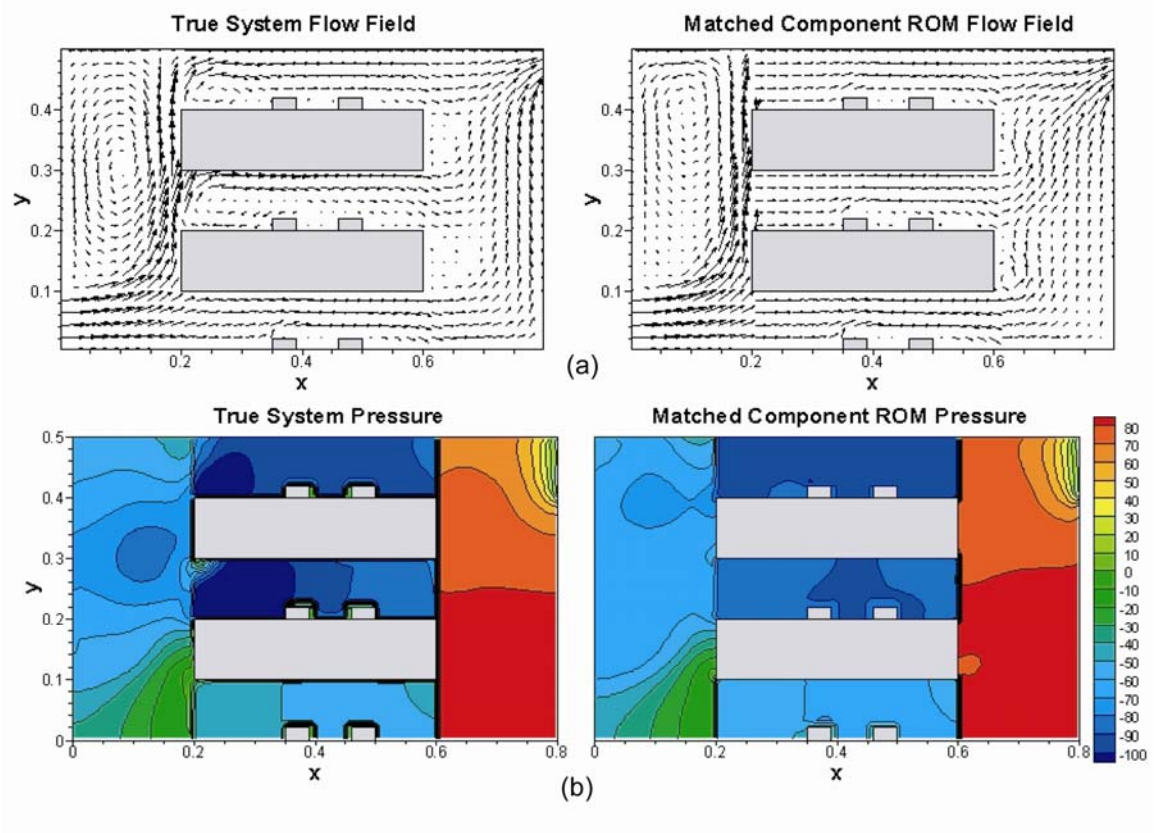


Figure 6.8. Exact and component-matched velocity and pressure fields [Pa]

The regions of greatest error occur at the interfaces where there is severe misalignment between the component velocity fields. The flow at the entrance to the uppermost server ($G_3^1 \rightarrow G_1^4$), plotted in Figure 6.9a, is almost entirely in the vertical direction which leads to a discontinuity in the approximate solution because the server ROM has been parametrized in terms of inlet mass flux only and has assumed a uniform inlet velocity with no vertical component. The lumped resistance only acts to attenuate the x-velocity

component and does not affect the y-velocity component. The recirculation region near the inlet to domains Ω^3 and Ω^4 is not resolved by the approximate solution only because the server ROM did not contain this behavior. This is *not* an inherent limit of the methodology presented here because, as with any methodology that couples component models, the interconnected models must contain the behavior passed to them from other components. Figure 6.9b compares the flow at the exhaust of a server (the Ω^4 – Ω^5 interface). The predictions are much more accurate and resolve the dominant flow features because the velocity from Ω^4 has a significant component normal to Γ_3^2 and the ROM for Ω^5 was constructed assuming a uniform inflow normal to the surface Γ_2^5.

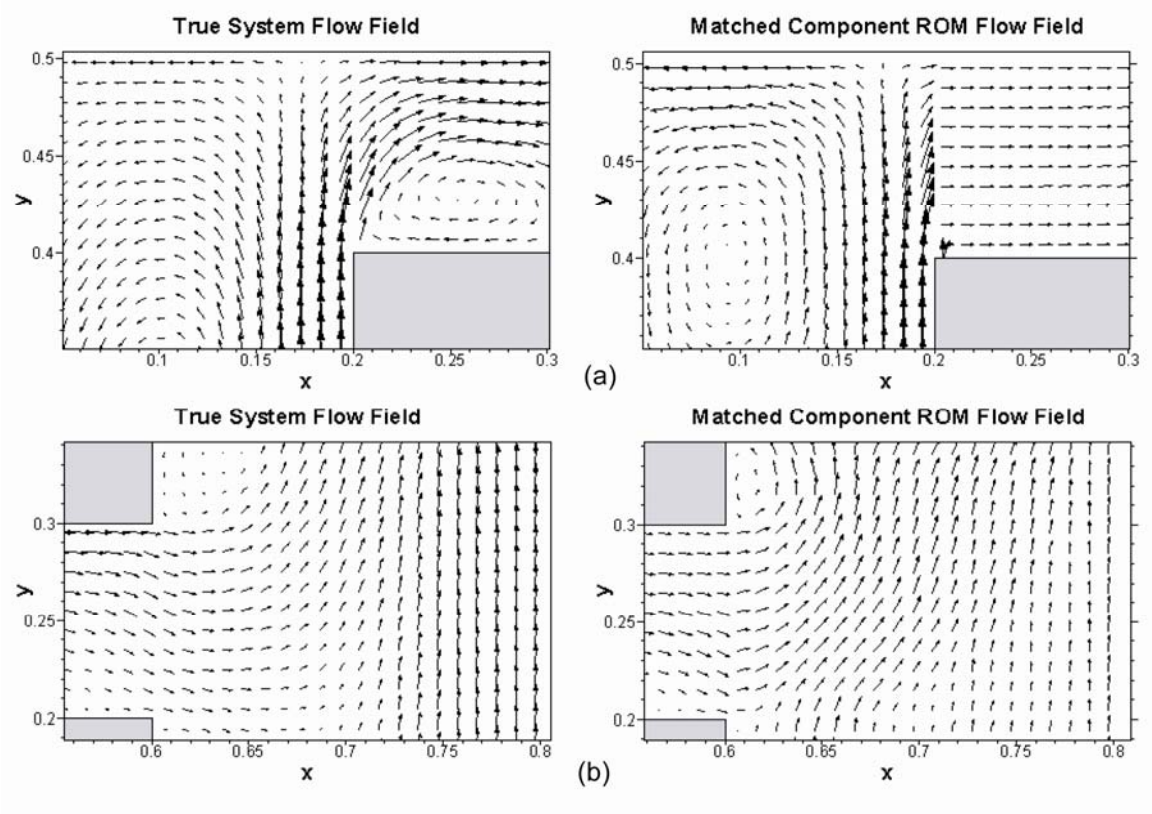


Figure 6.9. Local velocity mismatch at interface a) Ω^1 – Ω^4 and b) Ω^3 – Ω^5, n.b. the vectors have been interpolated to a coarse grid for illustrative purposes

The error of the assembled system can be attributed to the ROM mass flux and pressure field approximations. The ROMs presented here satisfy the target mass flow rate to ~3% and the pressure field approximation is accurate point-wise on the control surfaces to ~5%. These error are coupled in the FNM procedure during the pressure correction step when the nodal pressures are updated as a function of the nodal mass imbalance. Another concern in connecting component models together is the propagation of error and the compounding of errors as more components are added to the system. POD-based ROMs manage this error because the individual models satisfy overall mass and energy balances, making the $O(10^{-2})$ variations between desired and ROM mass and energy fluxes the principal error contribution.

6.5. Additional Component Parametrization

The discontinuity in the velocity field at the interface region is not an inherent limitation in assembling POD-based ROMs, rather is an issue pertaining to all interconnected systems modeling approaches. One such example occurs in block diagram modeling of dynamic systems. Suppose one component model has its behavior tuned to the frequency range $5 \leq f \leq 15$ Hz and its output is connected to a model that has a resonance at 25 Hz. The full-scale system will never see the effects of that resonance because it cannot be exited by the upstream model. The recirculation region not resolved by the server ROM in Figure 6a is caused by the same effect; the upstream plenum model introduces a non-normal velocity component which the server ROM cannot account for. This poses no problem in assembling the components since the matching conditions of

mass flux and pressure are still available, it just results in an unphysical discontinuity in the fully-assembled velocity field.

The server ROM was parameterized in terms of inlet mass flux only, but had velocity inlet direction been accounted for the parameterization, the server model would produce much more accurate velocity field predictions near the interface. An updated server ROM is developed from modes parametrized by the inlet mass flux and average y-velocity on inlet, i.e. $F = f(G,\bar{v})$. Incorporating this model into the full-scale system shows that the recirculation region can be resolved in the server ROM, see Figure 6.10. The pressure field is nearly identical to that of Figure 6.8 except for small regions in the upper portion of the interface between the intake plenum and servers 2 and 3 where the inlet velocity impinges on the upper wall.

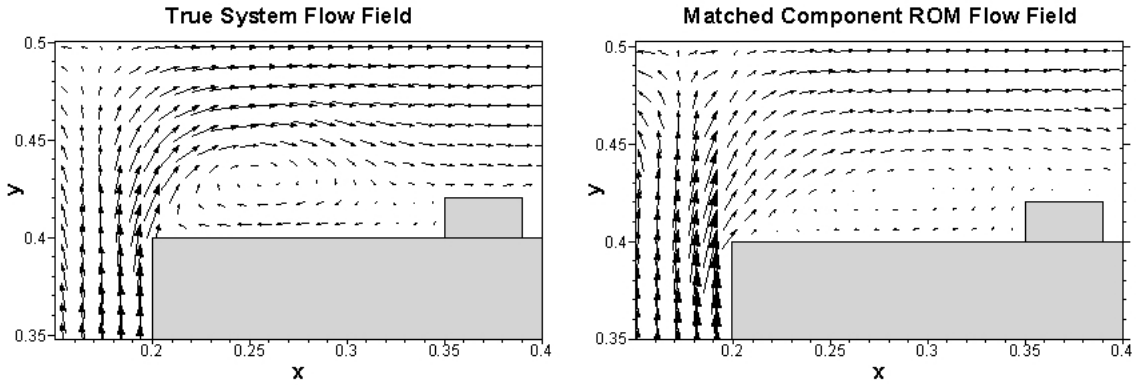


Figure 6.10. $\Omega^1 - \Omega^4$ interface for updated server ROM

The full-scale model presented here illustrates a number of issues concerning the interconnection of full-field models. The main issue concerns the severe misalignment of the velocity field from either side of the interface between two component models, which can be resolved by additional parametrization of the component ROM. This example is

especially difficult because the lumped parameter model does not act as a flow straightener by attenuating the velocity component parallel to the interface. Even with such limited ROM model behavior, the result systems-level model produces reasonably accurate representation of the flow and pressure fields.

7. CONCLUDING REMARKS

The reduced-order modeling framework developed here presents an accurate methodology for full-field approximations of steady turbulent flows and turbulent convection. The methodology presented here is equally applicable to experimental data and could be used to create low-dimensional models of stochastically characterized systems and integrated into large-scale simulations, creating an experimentally validated, computationally efficient modeling methodology for complex systems. Slightly less accurate models may be far superior to large expensive models during early system design and optimization, where many different parameter values and component interactions may need to be evaluated. The low-dimensional framework developed here also has the advantage of characterizing distributed parameter systems in state space using integral conditions, alleviating the need to specify detailed flow and heat transfer profiles that are often unknown. It also does not require the evaluation of the governing equations, making it well suited for inverse problems and parameter identification studies.

Reduced-order models can be used to efficiently bridge length scales in modeling and analysis of complex thermal – fluids systems. Meshes constructed for CFD/HT simulations of systems-level electronics thermal management must adequately resolve sharp gradients at the smallest modeled length scale, often producing excessively large models. Conventional CFD/HT techniques require a continuous computational grid that causes the smallest mesh features to dictate the overall grid size. Computational limitations often limit the smallest features than can be accurately modeled. Reduced-

order models (ROMs) based on the framework presented here can be integrated into large-scale CFD/HT simulations to construct efficient models across various length scales. These models accurately couple the data present on the CFD/HT grid and produce full-field approximations at length scales below the computational grid size without requiring the smallest features in the simulation to determine the full model size. The true efficiency in the reduced-order modeling framework is the ability to assemble the models as various components in a full-scale model.

7.1. Summary

The traditional proper orthogonal decomposition (POD) methodology has been reformulated to treat steady parametric turbulent flows and forced convective flows for a predefined range of boundary conditions. Orthogonal complement POD subspaces were introduced to treat inhomogeneous boundary conditions, eliminating the additional effort required by homogenization procedures and extending the reduced-order methodology to a wide range of flow parameters. A flux matching procedure (FMP) was formulated to evaluate the modal weight coefficients after the standard methods of Galerkin projection were shown to be ineffective for developing parametric reduced-order models. The integral conditions of the FMP are used in a state space residual expansion to mimic the POD subspace eigenvalue spectra and converge toward the desired parametric conditions. An implicit coupling procedure was developed to link the temperature and velocity fields, greatly improving the accuracy of low-dimensional temperature predictions. The overall reduced-order modeling framework presented here was able to reduce numerical models

containing 10^3 - 10^5 DOF down to less than 20 DOF for an order 10^2 – 10^3 reduction, while still retaining greater that 90% accuracy over the domain.

Rigorous *a posteriori* error bounds were developed by partitioning the error into in-plane and out of plane components with special attention to the error associated with using the FMP to evaluate the modal weight coefficients. Dual residual methods were used to show that the flux matching procedure converges and is computationally superior approach for low-dimensional modeling of steady turbulent flows and convection.

Parametric reduced-order models were then used as component-level models and assembled together to model a full-scale system. Distributed parameter ROMs were combined with lumped parameter models of fluid moving devices and flow resistances to model a pressure-driven system. Accurate approximations of the component pressure fields were constructed using response surface techniques, with the pressure becoming the driving force between the different sub-domains. Flow network modeling was used as the component handshaking procedure to couple the component ROMs through the interfacial mass flux.

The methodology to construct POD-based ROMs for parametric flows is schematically summarized in Figure 7.1 below, which segregates the methodology into model development, single component ROM solution and the interconnection of ROMs for pressure-driven flows. The model development consists of constructing a numerical or experimental representation of the physical system to be modeled and then observing the system under various parameter values to form the observation database. To use the ROM to construct a new approximate solution, the desired approximation is expressed in terms of the model parameters, since this is the only known portion of the solution. The

reduced-order solution begins with the selection of an observation to serve as the source function. Orthogonal complement subspaces POD subspaces are then be constructed and the flux matching procedure is used to evaluate the modal weight coefficients. The interconnection of ROMs to model systems-level flows requires the computation of a pressure response surface in the model development portion of the process. Additional lumped parameter models are specified and the component interfacial mass fluxes are computed using the standard SIMPLE algorithm. The solution procedure to solve for the individual component velocity and pressure fields is repeatedly used in connecting ROMs because the SIMPLE procedure requires the solution for each POD-based ROM during each iteration.

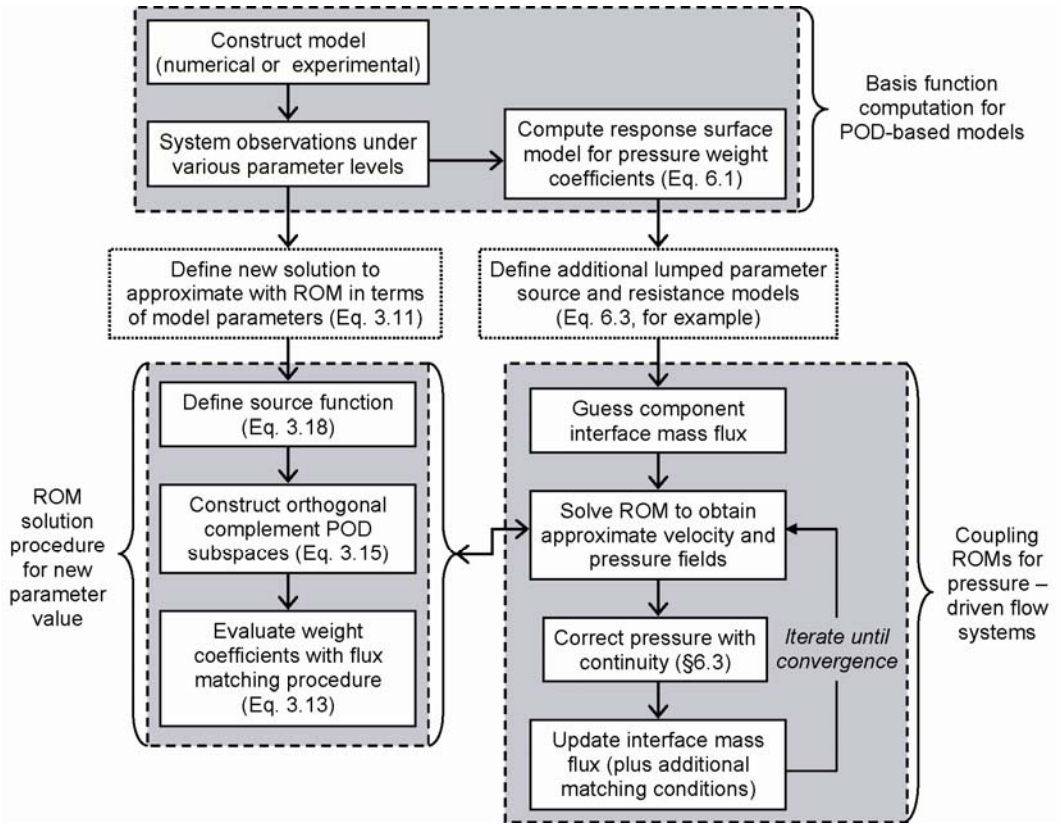


Figure 7.1. Reduced-order model (ROM) methodology flowchart

7.2. Future Work

The low-dimensional modeling methodology has focused on transport processes for electronics cooling, especially internal turbulent convection with conjugate conduction. However, the framework is general enough to accommodate a wide range of other thermal – fluid mechanisms including radiation, electrohydrodynamics and chemically reacting flows. This procedure can also be extended to non-continuum simulations such as molecular dynamics with varying parameters or boundary conditions to integrate micro- and nanoscale mass, momentum and energy transfers into full-scale engineering computations. This would allow one to efficiently link non-continuum models into complex engineering systems, without requiring mathematical homogenization procedures.

The development of reduced-order models represents a paradigm shift in computational thermal sciences. Much research in this area is currently focused on higher order approximation schemes and advanced mesh generation, intrinsically enlarging the computational effort. Given the limitations in modeling complex geometries and nonlinear phenomena, model reduction provides a tool to rapidly characterize the dominant behavior of the system, which can be used in conjunction with robust design techniques to perform optimization studies and design control schemes for systems with modeling and operational uncertainty. The reduced-order modeling framework was developed for use in conjunction with design and optimization tools. This research bridges the gap between the large data sets generated by experimental and numerical solutions of thermal – fluid transport in complex geometries, and forms a link to research in design. Low-dimensional modeling can also be incorporated into the

emerging field of simulation-based design, where experimentally validated, reduced-order models can be used to efficiently analyze and design thermal-fluid systems involving various transport mechanisms across several decades of length and time scales.

REFERENCES

[1] Patel, C. D., Bash, C. E., Belady, C., Stahl, L., and Sullivan, D., 2001, "Computational Fluid Dynamics Modeling of High Compute Density Data Centers to Assure System Inlet Air Specifications", *Proc. of Proc. IPACK'01 - The Pacific Rim / ASME International Electronics Packaging Technical Conference and Exhibition* Kauai, HI.

[2] "Heat Density Trends in Data Processing, Computer Systems and Telecommunications Equipment", 2000, The Uptime Institute, available at: http://www.upsite.com/TUIpages/tuiwhite.html, accessed: March 24, 2006.

[3] ASHRAE, 2005, *Datacom Equipment Power Trends and Cooling Applications*, American Society of Heating, Refrigeration and Air-Conditioning Engineers, Technical Committee TC9.9, Atlanta, GA.

[4] "Data Center Energy Benchmarking Case Study", 2003, Lawrence Berkeley National Laboratory and Rumsey Engineers, Inc., available at: http://datacenters.lbl.gov/, accessed: March 24, 2006.

[5] Sullivan, R. F., 2002, "Alternating Cold and Hot Aisles Provides More Reliable Cooling for Server Farms", The Uptime Institute, available at: http://www.upsite.com/TUIpages/tuiwhite.html, accessed: March 24, 2006.

[6] Boucher, T. D., Auslander, D. M., Bash, C. E., Federspiel, C. C., and Patel, C. D., 2004, "Viability of Dynamic Cooling Control in a Data Center Environment", *Proc. of ITHERM 2004 - Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, Las Vegas, NV.

[7] Schmidt, R. R., Karki, K. C., Kelkar, K. M., Radmehr, A., and Patankar, S. V., 2001, "Measurements and Predictions of the Flow Distribution Through Perforated Tiles in Raised Floor Data Centers", *Proc. of IPACK'01 - The Pacific Rim / ASME International Electronics Packaging Technical Conference and Exhibition*, Kauai, HI.

[8] Schmidt, R. R. and Shaukatullah, H., 2002, "Computer and Telecommunications Equipment Room Cooling: A Review of the Literature", *Proc. of ITHERM 2002 - Eight Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, San Diego, CA.

[9] Schmidt, R. R., Karki, K. C., and Patankar, S. V., 2004, "Raised-Floor Data Center: Perforated Tile Flow Rates for Various Tile Layouts", *Proc. of ITHERM 2004 - Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, Las Vegas, NV, USA.

[10] Radmehr, A., Schmidt, R. R., Karki, K. C., and Patankar, S. V., 2005, "Distributed Leakage Flow in Raised-Floor Data Centers", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Franciso, CA.

[11] Van Gilder, J. W. and Schmidt, R. R., 2005, "Airflow Uniformity Through Perforated Tiles in a Rasied-Floor Data Center", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Francisco, CA.

[12] Patel, C. D., Sharma, R., Bash, C. E., and Beitelmal, A., 2002, "Thermal Considerations in Cooling of Large Scale High Compute Density Data Centers", *Proc. of ITHERM 2002 - Eight Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, San Diego, CA.

[13] Schmidt, R. R. and Cruz, E., 2002, "Raised-Floor Compouter Data Center: Effect on Rack Inlet Temperatures of Chilled Air Exiting form Both the Hot and Cold Aisles", *Proc. of ITHERM 2002 - Eight Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, San Diego, CA.

[14] Schmidt, R. R. and Cruz, E., 2003, "Cluster of High Powered Racks within a Raised Floor Computer Data Center: Effects of Perforated Tiles Flow Distribution on Rack Inlet Air Temperature", *Proc. of IMECE'03 – ASME International Mechanical Engineering Congress and R&D Exposition, Washington D.C*,

[15] Schmidt, R. R. and Iyengar, M., 2005, "Effect of Data Center Layout on Rack Inlet Air Temperatures", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Francisco, CA.

[16] Bhopte, S., Agonafer, D., Schmidt, R. R., and Sammakia, B., 2005, "Optimization of Data Center Room Layout to Minimize Rack Inlet Air Temperature", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Francisco, CA.

[17] Rambo, J. and Joshi, Y., 2003, "Physical Models in Data Center Airflow Simulations", *Proc. of IMECE'03 – ASME International Mechanical Engineering Congress and R&D Exposition*, Washington D.C.

[18] Iyengar, M., Schmidt, R. R., Sharma, A., McVicker, G., Shrivastava, S., Sri-Jayantha, S., Amemiya, Y., Dang, H., Chainer, T., and Sammakia, B., 2005, "Thermal Characterization of Non-Raised Floor Air Cooled Data Centers Using Numerical Modeling", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Francisco, CA.

[19] Shrivastava, S., Sammakia, B., Schmidt, R. R., and Iyengar, M., 2005, "Comparative Analysis of Different Data Center Airflow Management Configurations", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Francisco, CA.

[20] Rambo, J. and Joshi, Y., 2003, "Multi-Scale Modeling of High Power Density Data Centers", *Proc. of IPACK'03 - The Pacific Rim / ASME International Electronics Packaging Technical Conference and Exhibition*, Kauai, HI.

[21] Rambo, J. and Joshi, Y. 2006, "Thermal Modeling of Technology Infrastructure Facilities: A Case Study of Data Centers", in: *The Handbook of Numerical Heat Transfer: Vol II*, W.J. Minkowycz, E.M. Sparrow, and J.Y. Murthy (Eds.), Taylor and Francis, New York, pp.

[22] Rambo, J. and Joshi, Y., *accepted for publication* 2006, "Convective Transport Processes in Data Centers", Numerical Heat Transfer A - Applications, pp.

[23] Sharma, R. K., Bash, C. E., and Patel, C. D., 2002, "Dimensionless Parameters for the Evaluation of Thermal Design and Performance of Large-Scale Data Centers", AIAA, **AIAA-2002-3091** pp.

[24] Sharma, R., Bash, C. E., Patel, C. D., and Beitelmal, M., 2004, "Experimental Investigation of Design and Performance of Data Centers", *Proc. of ITHERM 2004 - Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, Las Vegas, NV.

[25] Norota, M., Hayama, H., Enai, M., Mori, T., and Kishita, M., 2003, "Research on Efficiency of Air Conditioning System for Data Center", *Proc. of INTELEC'03 - 25th International Telecommunications Energy Conference,*, Yokohama, Japan.

[26] Shah, A., Carey, V. P., Bash, C. E., and Patel, C. D., 2003, "Exergy Analysis of Data Center Thermal Management Systems", *Proc. of IMECE'03 – ASME International Mechanical Engineering Congress and R&D Exposition*, Washington, D.C.

[27] Shah, A. J., Carey, V. P., Bash, C. E., and Patel, C. D., 2005, "Exergy-Based Optimization Strategies for Multi-Component Data Center Thermal Management: Part I: Analysis", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Francisco, CA.

[28] Shah, A. J., Carey, V. P., Bash, C. E., and Patel, C. D., 2005, "Exergy-Based Optimization Strategies for Multi-Component Data Center Thermal Management: Part II: Application and Validation", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Francisco, CA.

[29] Rambo, J. and Joshi, Y., 2005, "Thermal Performance Metrics for Arranging Forced Air Cooled Servers in a Data Processing Cabinet", ASME Journal of Electronic Packaging, **127** (4), pp. 452-459.

[30] Pope, S. B., 2000, *Turbulent Flows*, Cambridge University Press, New York.

[31] Ma, X., Karniadakis, G. E., Park, H. M., and Gharib, M., 2002, "DPIV/T-Driven Convective Heat Transfer Simulation", International Journal of Heat and Mass Transfer, **45** pp. 3517-3527.

[32] Shapiro, B., 2003, "Creating Compact Models for Electronic Systems: An Overview and Suggested Use of Existing Model Reduction and Experimental System Identification Tools", IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part A, **26** (1), pp. 165-172.

[33] Bar-Cohen, A. and Krueger, W., 1997, "Thermal Characterization of Chip Packages - Evolutionary Design of Compact Models", IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part A, **20** (4), pp. 399-410.

[34] Lasance, C. J. M., 2003, "Special Section on Compact Thermal Modeling", IEEE Transactions on Components and Packaging Technologies, **26** (1), pp. 134-135.

[35] Lasance, C. J. M., 2004, "Highlights from the European Thermal Project PROFIT", ASME Journal of Electronic Packaging, **126** pp. 565-570.

[36] Lasance, C. J. M., Rosten, H. I., and Parry, J. D., 1997, "The World of Thermal Characterization According to DELPHI - Part II: Experimental and Numerical Methods", IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part A, **20** (4), pp. 392-398.

[37] Rosten, H. I., Lasance, C. J. M., and Parry, J. D., 1997, "The World of Thermal Characterization According to DELPHI - Part I: Background to DELPHI", IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part A, **20** (4), pp. 384-391.

[38] Holmes, P., Lumley, J. L., and Berkooz, G., 1996, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Great Britain.

[39] Deane, A. E., Kevrekidis, I. G., Karniadakis, G. E., and Orszag, S. A., 1991, "Low-Dimensional Models for Complex Geometry Flows: Application to Grooved Channels and Circular Cylinders", Physics of Fluids A, **3** (10), pp. 2337-2354.

[40] Park, H. M. and Kim, O. Y., 2001, "Reduction of Modes for the Control of Viscous Flows", International Journal of Engineering Science, **39** pp. 177-200.

[41] Ravindran, S. S., 2002, "Control of Flow Separation Over a Forward-Facing Step by Model Reduction", Computer Methods in Applied Mechanics and Engineering, **191** pp. 1924-1942.

[42] Ravindran, S. S., 2002, "Adaptive Reduced-Order Controllers for a Thermal Flow Using Proper Orthogonal Decomposition", SIAM Journal of Scientific Computing, **23** (6), pp. 1924-1942.

[43] Ma, X. and Karniadakis, G. E., 2002, "A Low-Dimensional Model for Simulating Three-Dimensional Cylinder Flows", Journal of Fluid Mechanics, **458** pp. 181-190.

[44] Park, H. M. and Cho, D. H., 1996, "The Use of the Karhunen-Loeve Decomposition for the Modeling of Distributed Parameter Systems", Chemical Engineering Science, **51** (1), pp. 81-98.

[45] Park, H. M. and Cho, D. H., 1996, "Low Dimensional Modeling of Flow Reactors", International Journal of Heat and Mass Transfer, **39** (16), pp. 3311-3323.

[46] Sirovich, L. and Park, H. M., 1990, "Turbulent Thermal Convection in a Finite Domain: Part I. Theory", Physics of Fluids, **2** (9), pp. 1649-1657.

[47] Sirovich, L. and Park, H. M., 1990, "Turbulent Thermal Convection in a Finite Domain: Part II. Numerical Results", Physics of Fluids, **2** (9), pp. 1649-1657.

[48] Tarman, I. H. and Sirovich, L., 1998, "Extensions of Karhunen-Loeve Based Approximations of Complicated Phenomena", Computer Methods in Applied Mechanics and Engineering, **155** pp. 359-368.

[49] Park, H. M. and Li, W. J., 2002, "Boundary Optimal Control of Natural Convection by Means of Mode Reduction", Journal of Dynamic Systems, Measurement and Control, **124** pp. 47-54.

[50] Sirisup, S. and Karniadakis, G. E., 2005, "Stability and Accuracy of Periodic Flow Solutions Obtained by a POD-Penalty Method", Physica D, **202** pp. 218-237.

[51] Taylor, J. A. and Glauser, M. N., 2004, "Towards Practical Flow Sensing and Control via POD and LSE Based Low-Dimensional Tools", ASME Journal of Fluids Engineering, **126** pp. 337-345.

[52] Utturkar, Y., Zhang, B., and Shyy, W., 2005, "Reduced-Order Description of Fluid Flow with Moving Boundaries by Proper Orthogonal Decomposition", International Journal of Heat and Fluid Flow, **26** pp. 276-288.

[53] Galletti, B., Bruneau, C. H., Zannetti, L., and Iollo, A., 2004, "Low-Order Modeling of Laminar Flow Regimes Past a Confined Square Cylinder", Journal of Fluid Mechanics, **503** pp. 161-170.

[54] Ly, H. V. and Tran, H. T., 2001, "Modeling and Control of Physical Processes Using Proper Orthogonal Decomposition", Mathematical and Computer Modeling, **33** pp. 223-236.

[55] Rempfer, D., 2000, "On Low-Dimensional Galerkin Models for Fluid Flow", Theoretical and Computational Fluid Dynamics, **14** (2), pp. 75-88.

[56] Christensen, E. A., Brons, M., and Sorensen, J. N., 2000, "Evaluation of Proper Orthogonal Decomposition - Based Techniques Applied to Parameter Dependent Nonturbulent Flows", SIAM Journal of Scientific Computing, **21** (4), pp. 1419-1434.

[57] Graham, M. D. and Kevrekidis, I. G., 1996, "Alternative Approaches to the Karhunen-Loeve Decomposition for Model Reduction and Data Compression", Computers in Chemical Engineering, **20** (5), pp. 495-506.

[58] Park, H. M. and Jung, W. S., 2001, "The Karhunen-Loeve Galerkin Method for the Inverse Natural Convection Problems", International Journal of Heat and Mass Transfer, **44** (1), pp. 155-167.

[59] Rowley, C. W., Colonius, T., and Murray, R. M., 2004, "Model Reduction for Compressible Flows Using POD and Galerkin Projection", Physica D, **189** pp. 115-129.

[60] Sirovich, L., 1987, "Turbulence and the Dynamics of Coherent Structures, Parts I - III", Quarterly of Applied Mathematics, **XLV** (3), pp. 561-590.

[61] Gottlieb, D. and Orszag, S. A., 1977, *Numerical Analysis of Spectral Methods*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

[62] Penrose, R., 1955, "A Generalized Inverse for Matrices", Proceedings of the Cambridge Philosophical Society, **51** pp. 406-413.

[63] Rathnam, M. and Petzold, L. R., 2003, "A New Look at Proper Orthogonal Decomposition", SIAM Journal of Numerical Analysis, **41** (5), pp. 1893-1925.

[64] Fluent Inc., 2001, *Fluent v. 6.1 Users Manual*, Fluent Incorporated, Lebanon, New Hampshire.

[65] Rolander, N., Rambo, J., Joshi, Y., and Mistree, F., 2005, "Towards Sustainable Design of Data Centers: Addressing the Lifecycle Mismatch Problem", *Proc. of IPACK'05 - International Electronic Packaging Technical Conference and Exhibition*, San Franciso, CA.

[66] Rolander, N., Rambo, J., Joshi, Y., Mistree, F., and Allen, J. K., *accepted for publication* 2006, "Robust Design of Turbulent Convective Systems Using the Proper Orthogonal Decomposition", ASME Journal of Mechanical Design: Special Issue Robust and Risk Based Design, pp.

[67] Haider, S. I., Rambo, J., and Joshi, Y., 2005, *Development of a Modular Framework for the System-level Thermal Modeling of the Shipboard Power-Electronics Cooling System* ONR-Contract 04-1-0335, monitored by Dr. Mark Spector, Georgia Institute of Technology, Atlanta, GA.

[68] Yoo, S.-Y., Park, J.-H., and Chung, M.-H., 2003, "Local Heat Transfer Characteristics in Simulated Electronic Modules", ASME Journal of Electronic Packaging, **125** pp. 362-368.

[69] Kader, B., 1981, "Temperature and Concentration Profiles in Fully Turbulent Boundary Layers", International Journal of Heat and Mass Transfer, **24** (9), pp. 1541-1544.

[70] Chen, Y.-M. and Wang, K.-C., 1998, "Experimental Study on the Forced Convection Flow in a Channel with Heated Blocks in Tandem", Experimental Thermal and Fluid Science, **16** pp. 286-298.

[71] Jones, W. P. and Launder, B. E., 1973, "The Calculation of Low-Reynolds-Number Phenomena with a Two-Equation Model of Turbulence", International Journal of Heat and Mass Transfer, **16** pp. 1119-1130.

[72] Rediniotis, O. K., Ko, J., and Kurdilla, A., 2002, "Reduced Order Nonlinear Navier-Stokes Models for Synthetic Jets", Journal of Fluids Engineering, **124** pp. 433-443.

[73] Rambo, J. and Joshi, Y., 2005, "Reduced Order Modeling of Steady Turbulent Flows Using the POD", *Proc. of ASME Summer Heat Transfer Conference*, San Francisco, CA.

[74] Meyer, M. and Matthies, H. G., 2003, "Efficient Model Reduction Non-Linear Dynamics Using the Karhunen-Loeve Expansion and Dual-Weighted-Residual Methods", Computational Mechanics, **31** pp. 179-191.

[75] Belady, C., Kelkar, K. M., and Patankar, S. V., 1995, "Improving Productivity in Electronic Packaging with Flow Network Modeling (FNM)", Electronics Cooling, **5** (1), pp. 36-40.

[76] Patankar, S. V., 1980, *Numerical Heat Transfer and Fluid Flow*, McGraw Hill, New York.